

A Multiscale Pyramid Transform for Graph Signals

David I Shuman, Mohammad Javad Faraji, and Pierre Vandergheynst

Abstract—Multiscale transforms designed to process analog and discrete-time signals and images cannot be directly applied to analyze high-dimensional data residing on the vertices of a weighted graph, as they do not capture the intrinsic geometric structure of the underlying graph data domain. In this paper, we adapt the Laplacian pyramid transform for signals on Euclidean domains so that it can be used to analyze high-dimensional data residing on the vertices of a weighted graph. Our approach is to study existing methods and develop new methods for the four fundamental operations of graph downsampling, graph reduction, and filtering and interpolation of signals on graphs. Equipped with appropriate notions of these operations, we leverage the basic multiscale constructs and intuitions from classical signal processing to generate a transform that yields both a multiresolution of graphs and an associated multiresolution of a graph signal on the underlying sequence of graphs.

Index Terms—Signal processing on graphs, multiresolution, spectral graph theory, graph downsampling, Kron reduction, spectral sparsification, Laplacian pyramid, interpolation

I. INTRODUCTION

Multiscale transform methods can reveal structural information about signals, such as singularities or irregular structural patterns, at different resolution levels. At the same time, via coarse-to-fine analysis, they provide a way to reduce the complexity and dimensionality of many signal processing tasks, often resulting in fast algorithms. However, multiscale transforms such as wavelets and filter banks designed to process analog and discrete-time signals and images cannot be directly applied to analyze high-dimensional data residing on the vertices of a weighted graph, as they do not capture the intrinsic geometric structure of the underlying graph data domain (see [2] for an overview of the main challenges of this emerging field of signal processing on graphs).

To address this issue, classical wavelets have recently been generalized to the graph setting in a number of different ways. Reference [2] contains a more thorough review of these graph wavelet constructions, which include, e.g., spatially-designed graph wavelets [3], diffusion wavelets [4], spectral graph wavelets [5], lifting based wavelets [6], [7] critically sampled

two-channel wavelet filter banks [8], critically-sampled spline wavelet filter banks [9], and multiscale wavelets on balanced trees [10]. Multiresolutions of graphs also have a long history in computational science problems including graph clustering, numerical solvers for linear systems of equations (often arising from discretized differential equations), combinatorial optimization problems, and computational geometry (see, e.g., [11]–[14] and references therein). We discuss some of the related work from these fields in Sections III-E and IV-E.

In this paper, we present a modular framework for adapting Burt and Adelson’s Laplacian pyramid transform [15] to the graph setting. Our main contributions are to (1) survey different methods for and desirable properties of the four fundamental graph signal processing operations of graph downsampling, graph reduction, generalized filtering, and interpolation of graph signals (Sections III-V); (2) present new graph downsampling and reduction methods, including downsampling based on the polarity of the largest Laplacian eigenvector and Kron reduction followed by spectral sparsification; and (3) leverage these fundamental operations to construct a new multiscale pyramid transform that yields both a multiresolution of a graph and a multi-scale analysis of a signal residing on that graph (Section VI). We also discuss some implementation approximations and open issues in Section VII, as it is important that the computational complexity of the resulting multiscale transform scales well with the number of vertices and edges in the underlying graph.

II. SPECTRAL GRAPH THEORY NOTATION

We consider connected, loopless (no edge connecting a vertex to itself), undirected, weighted graphs. We represent such a graph by the triplet $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, w\}$, where \mathcal{V} is a set of N vertices, \mathcal{E} is a set of edges, and $w : \mathcal{E} \rightarrow \mathbb{R}^+$ is a weight function that assigns a non-negative weight to each edge. An equivalent representation is $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, where \mathbf{W} is a $N \times N$ weighted adjacency matrix with nonnegative entries

$$W_{ij} = \begin{cases} w(e), & \text{if } e \in \mathcal{E} \text{ connects vertices } i \text{ and } j \\ 0, & \text{if no edge connects vertices } i \text{ and } j \end{cases}.$$

In unweighted graphs, the entries of the adjacency matrix \mathbf{W} are ones and zeros, with a one corresponding to an edge between two vertices and a zero corresponding to no edge. The degree matrix \mathbf{D} is a diagonal matrix with an i^{th} diagonal element $D_{ii} = d_i = \sum_{j \in \mathcal{N}_i} W_{ij}$, where \mathcal{N}_i is the set of vertex i ’s neighbors in \mathcal{G} . Its maximum element is $d_{\max} := \max_{i \in \mathcal{V}} \{d_i\}$. We denote the combinatorial graph Laplacian by $\mathcal{L} := \mathbf{D} - \mathbf{W}$, the normalized graph Laplacian by $\tilde{\mathcal{L}} := \mathbf{D}^{-\frac{1}{2}} \mathcal{L} \mathbf{D}^{-\frac{1}{2}}$, and their respective eigenvalue and eigenvector pairs by $\{(\lambda_\ell, \mathbf{u}_\ell)\}_{\ell=0,1,\dots,N-1}$ and $\{(\tilde{\lambda}_\ell, \tilde{\mathbf{u}}_\ell)\}_{\ell=0,1,\dots,N-1}$. Then \mathbf{U} and $\tilde{\mathbf{U}}$ are the matrices whose columns are equal to the

This work was supported in part by FET-Open grant number 255931 UNLocX.

Part of this work appeared in the master’s thesis [1] and was presented at the Duke Workshop on Sensing and Analysis of High-Dimensional Data in July 2011. Slides and video are available at http://sahd.pratt.duke.edu/2011_files/Videos_and_Slides.html

David I Shuman is with the Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, MN 55105, USA (email: dshuman1@macalester.edu). Mohammad Javad Faraji is with the Computational Neuroscience Laboratory (LCN), Ecole Polytechnique Fédérale de Lausanne (EPFL), School of Computer and Communication Sciences and Brain Mind Institute, School of Life Sciences, CH-1015 Lausanne, Switzerland (email: mohammadjavad.faraji@epfl.ch). Pierre Vandergheynst is with the Signal Processing Laboratory (LTS2), Ecole Polytechnique Fédérale de Lausanne (EPFL), Institute of Electrical Engineering, CH-1015 Lausanne, Switzerland (email: pierre.vandergheynst@epfl.ch).

eigenvectors of \mathcal{L} and $\tilde{\mathcal{L}}$, respectively. We assume without loss of generality that the eigenvalues are monotonically ordered so that $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{N-1}$, and we denote the maximum eigenvalues and associated eigenvectors by $\lambda_{\max} = \lambda_{N-1}$ and $\mathbf{u}_{\max} = \mathbf{u}_{N-1}$. The maximum eigenvalue λ_{\max} is said to be *simple* if $\lambda_{N-1} > \lambda_{N-2}$.

III. GRAPH DOWNSAMPLING

Two key components of multiscale transforms for discrete-time signals are downsampling and upsampling.¹ To down-sample a discrete-time sample by a factor of two, we remove every other component of the signal. To extend many ideas from classical signal processing to the graph setting, we need to define a notion of downsampling for signals on graphs. Yet, it is not at all obvious what it means to remove *every other* component of a signal $\mathbf{f} \in \mathbb{R}^N$ defined on the vertices of a graph. In this section, we outline desired properties of a downsampling operator for graphs, and then go on to suggest one particular downsampling method.

Let $\mathcal{D} : \mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\} \rightarrow 2^{\mathcal{V}}$ be a graph downsampling operator that maps a weighted, undirected graph to a subset of vertices \mathcal{V}_1 to keep. The complement $\mathcal{V}_1^c := \mathcal{V} \setminus \mathcal{V}_1 = \{v \in \mathcal{V} : v \notin \mathcal{V}_1\}$ is the set of vertices that \mathcal{D} removes from \mathcal{V} . Ideally, we would like the graph downsampling operator \mathcal{D} to have the following properties:

- (D1) It removes approximately half of the vertices of the graph (or, equivalently, approximately half of the components of a signal on the vertices of the graph); i.e., $|\mathcal{D}(\mathcal{G})| = |\mathcal{V}_1| \approx \frac{|\mathcal{V}|}{2}$.
- (D2) The set of removed vertices are not connected with edges of high weight, and the set of kept vertices are not connected with edges of high weight; i.e., if $i, j \in \mathcal{V}_1$, then W_{ij} is low, and if $i, j \in \mathcal{V}_1^c$, then W_{ij} is low.
- (D3) It has a computationally efficient implementation.

A. Vertex Selection Using the Largest Eigenvector of the Graph Laplacian

The method we suggest to use for graph downsampling is to select the vertices to keep based on the polarity of the components of the largest eigenvector; namely, let

$$\mathcal{V}_1 = \mathcal{V}_+ := \{i \in \mathcal{V} : u_{\max}(i) \geq 0\}. \quad (1)$$

We refer to this method as the largest eigenvector vertex selection method. A few remarks are in order regarding this choice of downsampling operator. First, the polarity of the largest eigenvector splits the graph into two components. In this paper, we choose to keep the vertices in \mathcal{V}_+ , and eliminate the vertices in $\mathcal{V}_- := \{i \in \mathcal{V} : u_{\max}(i) < 0\}$, but we could just as easily do the reverse, or keep the vertices in $\mathcal{V}_{\text{big}} := \arg \max_{\mathcal{V}_1 \in \{\mathcal{V}_+, \mathcal{V}_-\}} |\mathcal{V}_1|$, for example. Second, for some graphs such as the complete graph, λ_{\max} is a repeated eigenvalue, so the polarity of \mathbf{u}_{\max} is not uniquely defined. In this case, we arbitrarily choose an eigenvector from the

eigenspace. Third, we could just as easily base the vertex selection on the polarity of the normalized graph Laplacian eigenvector, $\tilde{\mathbf{u}}_{\max}$ associated with the largest eigenvalue, $\tilde{\lambda}_{\max}$. In some cases, such as the bipartite graphs discussed next, doing so yields exactly the same selection of vertices as downsampling based on the largest non-normalized graph Laplacian eigenvector; however, this is not true in general.

In the following sections, we motivate the use of the largest eigenvector of the graph Laplacian from two different perspectives - first from a more intuitive view as a generalization of downsampling techniques for special types of graphs, and then from a more theoretical point of view by connecting the vertex selection problem to graph coloring, spectral clustering, and nodal domain theory.

B. Special Case: Bipartite Graphs

There is one situation in which there exists a fairly clear notion of removing every other component of a graph signal - when the underlying graph is *bipartite*. A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$ is *bipartite* if the set of vertices \mathcal{V} can be partitioned into two subsets \mathcal{V}_1 and \mathcal{V}_1^c so that every edge $e \in \mathcal{E}$ links one vertex in \mathcal{V}_1 with one vertex in \mathcal{V}_1^c . In this case, it is natural to downsample by keeping all of the vertices in one of the subsets, and eliminating all of the vertices in the other subset. In fact, as stated in the following theorem, the largest eigenvector downsampling method does precisely this in the case of bipartite graphs.

Theorem 1 (Roth, 1989): For a connected, bipartite graph $\mathcal{G} = \{\mathcal{V}_1 \cup \mathcal{V}_1^c, \mathcal{E}, \mathbf{W}\}$, the largest eigenvalues, λ_{\max} and $\tilde{\lambda}_{\max}$, of \mathcal{L} and $\tilde{\mathcal{L}}$, respectively, are simple, and $\tilde{\lambda}_{\max} = 2$. Moreover, the polarity of the components of the eigenvectors \mathbf{u}_{\max} and $\tilde{\mathbf{u}}_{\max}$ associated with λ_{\max} and $\tilde{\lambda}_{\max}$ both split \mathcal{V} into the bipartition \mathcal{V}_1 and \mathcal{V}_1^c . That is, for $\mathbf{v} = \mathbf{u}_{\max}$ or $\mathbf{v} = \tilde{\mathbf{u}}_{\max}$,

$$\begin{aligned} v(i)v(j) &> 0, \text{ if } i, j \in \mathcal{V}_1 \text{ or } i, j \in \mathcal{V}_1^c, \text{ and} \\ v(i)v(j) &< 0, \text{ if } i \in \mathcal{V}_1, j \in \mathcal{V}_1^c \text{ or } i \in \mathcal{V}_1^c, j \in \mathcal{V}_1. \end{aligned} \quad (2)$$

If, in addition, \mathcal{G} is k -regular ($d_i = k, \forall i \in \mathcal{V}$), then $\lambda_{\max} = 2k$, and

$$\mathbf{u}_{\max} = \tilde{\mathbf{u}}_{\max} = \begin{cases} \frac{1}{\sqrt{N}}, & \text{if } i \in \mathcal{V}_1 \\ -\frac{1}{\sqrt{N}}, & \text{if } i \in \mathcal{V}_1^c \end{cases}.$$

The majority of the statements in Theorem 1 follow from results of Roth in [16], which are also presented in [17, Chapter 3.6].

The path, ring (with an even number of vertices), and finite grid graphs, which are shown in Figure 1, are all examples of bipartite graphs and all have simple largest graph Laplacian eigenvalues. Using the largest eigenvector downsampling method leads to the elimination of every other vertex on the path and ring graphs, and to the quincunx sampling pattern on the finite grid graph (with or without boundary connections).

Trees (acyclic, connected graphs) are also bipartite. An example of a tree is shown in Figure 1(e). Fix an arbitrary vertex r to be the root of the tree \mathcal{T} , let \mathcal{Y}_r^0 be the singleton set containing the root, and then define the sets $\{\mathcal{Y}_r^t\}_{t=1,2,\dots}$ by $\mathcal{Y}_r^t := \{i \in \mathcal{V} : i \text{ is } t \text{ hops from the root vertex } r \text{ in } \mathcal{T}\}$. Then the polarity of the components of largest eigenvector

¹We focus here on downsampling, as we are only interested in upsampling previously downsampled graphs. As long as we track the positions of the removed components of the signal, it is straightforward to upsample by inserting zeros back into those components of the signal.

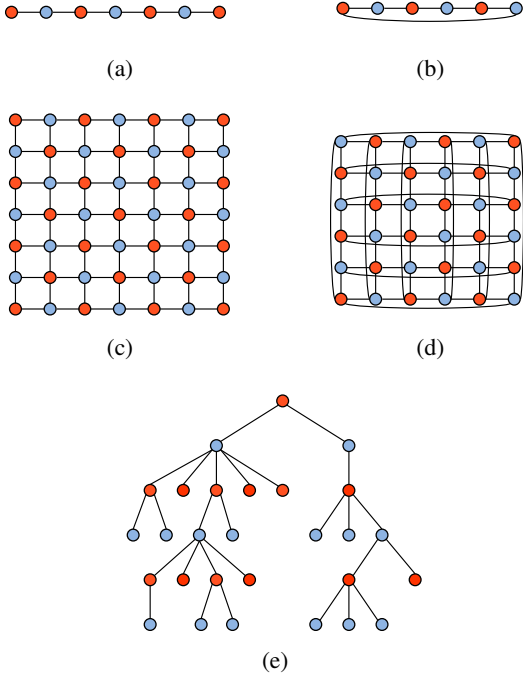


Fig. 1. Examples of partitioning structured graphs into two sets (red and blue) according to the polarity of the largest eigenvector of the graph Laplacian. For the path graph in (a) and the ring graph in (b), the method selects every other vertex like classical downsampling of discrete-time signals. For the finite grid graphs with the ends unconnected (c) and connected (d), the method results in the quincunx sampling pattern. For a tree graph in (e), the method groups the vertices at every other depth of the tree.

of the graph Laplacian splits the vertices of the tree into two sets according to the parity of the depths of the tree. That is, if we let $\mathcal{Y}_r^{\text{even}} := \cup_{t=0,2,\dots} \mathcal{Y}_r^t$ and $\mathcal{Y}_r^{\text{odd}} := \cup_{t=1,3,\dots} \mathcal{Y}_r^t$, then $\mathcal{Y}_r^{\text{even}} = \mathcal{V}_+$ and $\mathcal{Y}_r^{\text{odd}} = \mathcal{V}_-$, or vice versa.

In related work, [18] and [19] suggest to downsample bipartite graphs by keeping all of the vertices in one subset of the bipartition, and [20] suggests to downsample trees by (ND2) keeping vertices at every other depth of the tree. Therefore, the largest eigenvector downsampling method can be seen as a generalization of those approaches.

C. Connections with Graph Coloring and Spectral Clustering

A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$ is k -colorable if there exists a partition of \mathcal{V} into subsets $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_k$ such that if vertices $i, j \in \mathcal{V}$ are connected by an edge in \mathcal{E} , then i and j are in different subsets in the partition. The chromatic number χ of a graph \mathcal{G} is the smallest k such that \mathcal{G} is k -colorable. Thus, the chromatic number of a graph is equal to 2 if and only if the graph is bipartite.

As we have seen with the examples in the previous section, when a graph is bipartite, it is easy to decide how to split it into two sets for downsampling. When the chromatic number of a graph is greater than two, however, we are interested in finding an *approximate coloring* [21]; that is, a partition that has as few edges as possible that connect vertices in the same

subset.² As noted by [21], the approximate coloring problem is in some sense dual to the problem of spectral clustering (see, e.g. [22] and references therein).

Aspvall and Gilbert [21] suggest to construct an approximate 2-coloring of an unweighted graph according to the polarity of the eigenvector associated with the most negative eigenvalue of the adjacency matrix. For regular graphs, the eigenvector associated with the most negative eigenvalue of the adjacency matrix is the same as the largest graph Laplacian eigenvector, and so the method of [21] is equivalent to the largest Laplacian eigenvector method for that special case.

D. Connections with Nodal Domain Theory

A positive (negative) *strong nodal domain* of \mathbf{f} on \mathcal{G} is a maximally connected subgraph such that $f(i) > 0$ ($f(i) < 0$) for all vertices i in the subgraph [17, Chapter 3]. A positive (negative) *weak nodal domain* of \mathbf{f} on \mathcal{G} is a maximally connected subgraph such that $f(i) \geq 0$ ($f(i) \leq 0$) for all vertices i in the subgraph, with $f(i) \neq 0$ for at least one vertex i in the subgraph.

A graph downsampling can be viewed as an assignment of positive and negative signs to vertices, with positive signs assigned to the vertices that we keep, and negative signs to the vertices that we eliminate. The goal of having few edges within either the removed set or the kept set is closely related to the problem of maximizing the number of nodal domains of the downsampling. This is because maximizing the number of nodal domains leads to nodal domains with fewer vertices, which results in fewer edges connecting vertices within the removed and kept sets.

Next, we briefly mention some general bounds on the number of nodal domains of eigenvectors of graph Laplacians:

- (ND1) \mathbf{u}_ℓ has at most ℓ weak nodal domains and $\ell + s - 1$ strong nodal domains, where s is the multiplicity of λ_ℓ [23], [17, Theorem 3.1].
- (ND2) The largest eigenvector \mathbf{u}_{\max} has N strong and weak nodal domains if and only if \mathcal{G} is bipartite. Moreover, if \mathcal{H} is an induced bipartite subgraph of \mathcal{G} with the maximum number of vertices, then the number of vertices in \mathcal{H} is an upper bound on the number of strong nodal domains of any eigenvector of a generalized Laplacian of \mathcal{G} [17, Theorem 3.27].
- (ND3) If λ_ℓ is simple and $u_\ell(i) \neq 0, \forall i \in \mathcal{V}$, the number of nodal domains of \mathbf{u}_ℓ is greater than or equal to $\ell - r$, where r is the number of edges that need to be removed from the graph in order to turn it into a tree³ [24].

Note that while both the lower and upper bounds on the number of nodal domains of the eigenvectors of graph Laplacians are monotonic in the index of the eigenvalue, the actual number of nodal domains is not always monotonic in the index of the eigenvalue (see, e.g., [25, Figure 1] for an example where they are not monotonic). Therefore, for arbitrary graphs,

²In other contexts, the term *approximate coloring* is also used in reference to finding a proper k -coloring of a graph in polynomial time, such that k is as close as possible to the chromatic number of the graph.

³Berkolaiko proves this theorem for Schrödinger operators, which encompass generalized Laplacians of unweighted graphs.

it is not guaranteed that the largest eigenvector of the graph Laplacian has more nodal domains than the other eigenvectors. For specific graphs such as bipartite graphs, however, this is guaranteed to be the case, as can be seen from the above bounds and Theorem 1. Finally, in the case of repeated graph Laplacian eigenvalues, [17, Chapter 5] presents a hillclimbing algorithm to search for an associated eigenvector with a large number of nodal domains.

E. Alternative Graph Downsampling Methods

We briefly mention some alternative graph downsampling/partitioning methods:

- 1) As mentioned above, [21] partitions the graph based on the polarity of the eigenvector associated with the most negative eigenvalue of the adjacency matrix.
- 2) We can do a k -means clustering on \mathbf{u}_{\max} with $k = 2$ to separate the sets. A closely related alternative is to set a non-zero threshold in (1). A flexible choice of the threshold can ensure that $|\mathcal{V}_1| \approx \frac{N}{2}$.
- 3) In [8], Narang and Ortega downsample based on the solution to the weighted max-cut problem:

$$\arg \max_{\mathcal{V}_1} \left\{ \text{cut}(\mathcal{V}_1, \mathcal{V}_1^c) \right\} = \arg \max_{\mathcal{V}_1} \left\{ \sum_{i \in \mathcal{V}_1} \sum_{i' \in \mathcal{V}_1^c} W_{ii'} \right\}. \quad (\text{GR9})$$

This problem is NP-hard, so approximations must be used for large graphs.

- 4) Barnard and Simon [26] choose \mathcal{V}_1 to be a maximal independent set (no edge connects two vertices in \mathcal{V}_1 and every vertex in \mathcal{V}_1^c is connected to some vertex in \mathcal{V}_1). Maximal independent sets are not unique, but can be found via efficient greedy algorithms.
- 5) Ron et al. propose the SelectCoarseNodes algorithm [13, Algorithm 2], which leverages an algebraic distance measure to greedily select those nodes which have large future volume, roughly corresponding to those nodes strongly connected to vertices in the eliminated set.

Additionally, there are a number of graph coarsening methods where the reduced set of vertices is not necessarily a subset of the original set of vertices. We discuss those further in Section IV-E.

IV. GRAPH REDUCTION

So we now have a way to downsample a graph to a subset of the vertices. However, in order to implement any type of multiscale transform that will require filtering operations defined on the subset of selected vertices, we still need to define a method to reduce the graph Laplacian on the original set of vertices to a new graph Laplacian on the subset of selected vertices. We refer to this process as *graph reduction*. For the purpose of multiscale transforms, we would ideally like the graph reduction method to have some or all of the properties listed below. Note that the following represents one example of desiderata and is neither exhaustive nor definitive.

- (GR1) The reduction is in fact a graph Laplacian (i.e., a symmetric matrix with row sums equal to zero and nonpositive off-diagonal elements).

It preserves connectivity. That is, if the original graph is connected, then the reduced graph is also connected.

The spectrum of the reduced graph Laplacian is representative of and contained in the spectrum of the original graph Laplacian.

It preserves structural properties (e.g., if the original graph is a tree or bipartite, the reduced graph is accordingly a tree or bipartite).

(GR5) Edges in the original graph that connect vertices in the reduced vertex set are preserved in the reduced graph.

(GR6) It is invertible or partially invertible with some side information; i.e., from the reduced graph Laplacian and possibly some other stored information, we can recover the original graph Laplacian.

(GR7) It is computationally efficient to implement.

(GR8) It preserves graph sparsity. That is, the ratio of the number of edges over potential edges $\left(\frac{|\mathcal{E}|}{N(N-1)/2} \right)$ does not increase too much from the original graph to the reduced graph. In addition to preserving computational benefits associated with sparse matrix-vector multiplication, keeping this ratio low ensures that the reduced graph also captures local connectivity information and not just global information.

There is some meaningful correspondence between the eigenvectors of the reduced graph Laplacian and the original graph Laplacian; for example, when restricted to the kept set of vertices, the lowest (smoothest) eigenvectors of the original Laplacian are similar to or the same as the eigenvectors of the reduced Laplacian.

In the following, we discuss in detail one particular choice of graph reduction that satisfies some, but not all, of these properties. In Section IV-E, we briefly mention a few other graph reduction methods.

A. Kron Reduction

The starting point for the graph reduction method we consider here is Kron reduction (see [27] and references therein). We start with a weighted graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, its associated graph Laplacian \mathcal{L} , and a subset $\mathcal{V}_1 \subsetneq \mathcal{V}$ of vertices (the set selected by downsampling, for example) on which we want to form a reduced graph.⁴ The Kron reduction of \mathcal{L} is the Schur complement [28] of \mathcal{L} relative to $\mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1^c}$; that is, the Kron reduction of \mathcal{L} is given by

$$\mathcal{K}(\mathcal{L}, \mathcal{V}_1) := \mathcal{L}_{\mathcal{V}_1, \mathcal{V}_1} - \mathcal{L}_{\mathcal{V}_1, \mathcal{V}_1^c} \mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1^c}^{-1} \mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1}, \quad (3)$$

where $\mathcal{L}_{\mathcal{A}, \mathcal{B}}$ denotes the $|\mathcal{A}| \times |\mathcal{B}|$ submatrix consisting of all entries of \mathcal{L} whose row index is in \mathcal{A} and whose column index is in \mathcal{B} . We can also uniquely associate with $\mathcal{L}^{\text{Kron-reduced}} = \mathcal{K}(\mathcal{L}, \mathcal{V}_1)$ a reduced weighted graph $\mathcal{G}^{\text{Kron-reduced}} = \{\mathcal{V}_1, \mathcal{E}^{\text{Kron-reduced}}, \mathbf{W}^{\text{Kron-reduced}}\}$ by letting

$$W_{ij}^{\text{Kron-reduced}} = \begin{cases} -\mathcal{L}_{ij}^{\text{Kron-reduced}} & \text{if } i \neq j, \\ 0 & \text{o.w.} \end{cases}, \quad (4)$$

⁴We assume $|\mathcal{V}_1| \geq 2$.

and taking $\mathcal{E}^{Kron-reduced}$ to be the set of edges with non-zero weights in (4).⁵ An example of Kron reduction is shown in Figure 2.

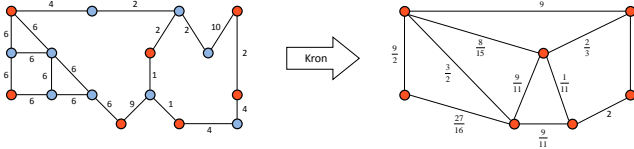


Fig. 2. An example of Kron reduction. The vertices in \mathcal{V}_1 are in red, and those in \mathcal{V}_1^c are in blue.

B. Properties of Kron Reduction

For a complete survey of the topological, algebraic, and spectral properties of the Kron reduction process, see [27]. In the next theorem, we summarize the results of [27] that are of most interest with regards to multiscale transforms.

Theorem 2 (Dörfler and Bullo, 2013): Let \mathcal{L} be a graph Laplacian of an undirected weighted graph \mathcal{G} on the vertex set \mathcal{V} , $\mathcal{V}_1 \subsetneq \mathcal{V}$ be a subset of vertices with $|\mathcal{V}_1| \geq 2$, $\mathcal{L}^{Kron-reduced} = \mathcal{K}(\mathcal{L}, \mathcal{V}_1)$ be the Kron-reduced graph Laplacian defined in (3), and $\mathcal{G}^{Kron-reduced}$ be the associated Kron-reduced graph. Then

- (K1) The Kron-reduced graph Laplacian in (3) is well-defined
- (K2) $\mathcal{L}^{Kron-reduced}$ is indeed a graph Laplacian
- (K3) If the original graph \mathcal{G} is connected (i.e., $\lambda_1(\mathcal{L}) > 0$), then the Kron-reduced graph $\mathcal{G}^{Kron-reduced}$ is also connected (i.e., $\lambda_1(\mathcal{L}^{Kron-reduced}) > 0$)
- (K4) If the original graph \mathcal{G} is loopless, then the Kron-reduced graph $\mathcal{G}^{Kron-reduced}$ is also loopless⁶
- (K5) Two vertices $i, j \in \mathcal{V}_1$ are connected by an edge if and only if there is a path between them in \mathcal{G} whose vertices all belong to $\{i, j\} \cup \mathcal{V}_1^c$
- (K6) Spectral interlacing: for all $\ell \in \{0, 1, \dots, |\mathcal{V}_1| - 1\}$,

$$\lambda_\ell(\mathcal{L}) \leq \lambda_\ell(\mathcal{L}^{Kron-reduced}) \leq \lambda_{\ell+|\mathcal{V}|-|\mathcal{V}_1|}(\mathcal{L}). \quad (5)$$

In particular, (5) and property (K4) above imply

$$\begin{aligned} 0 = \lambda_{\min}(\mathcal{L}) &= \lambda_{\min}(\mathcal{L}^{Kron-reduced}) \\ &\leq \lambda_{\max}(\mathcal{L}^{Kron-reduced}) \leq \lambda_{\max}(\mathcal{L}) \end{aligned} \quad (6)$$

- (K7) Monotonic increase of weights: for all $i, j \in \mathcal{V}_1$, $W_{ij}^{Kron-reduced} \geq W_{ij}$
- (K8) Resistance distance [29] preservation: for all $i, j \in \mathcal{V}_1$,

$$\begin{aligned} d_{R_G}(i, j) &:= (\delta_i - \delta_j)^T \mathcal{L}^\dagger (\delta_i - \delta_j) \\ &= (\delta_i - \delta_j)^T (\mathcal{L}^{Kron-reduced})^\dagger (\delta_i - \delta_j), \end{aligned}$$

where \mathcal{L}^\dagger is the pseudoinverse of \mathcal{L} and δ_i is the Kronecker delta (equal to 1 at vertex i and 0 at all other vertices)

⁵In (4), we still index the new weights by the same vertex indices i and j . In practice, the vertices i and j will receive new ordinal indices between 1 and $N^{Kron-reduced} = |\mathcal{V}_1|$, the number of vertices in the reduced graph.

⁶While [27] considers the more general case of *loopy Laplacians*, the graphs we are interested in are all *loopless*; that is, there are no self-loops connecting a vertex to itself, and therefore $\mathcal{L}_{nn} = \sum_{i \in \mathcal{V}, i \neq n} W_{ni}$.

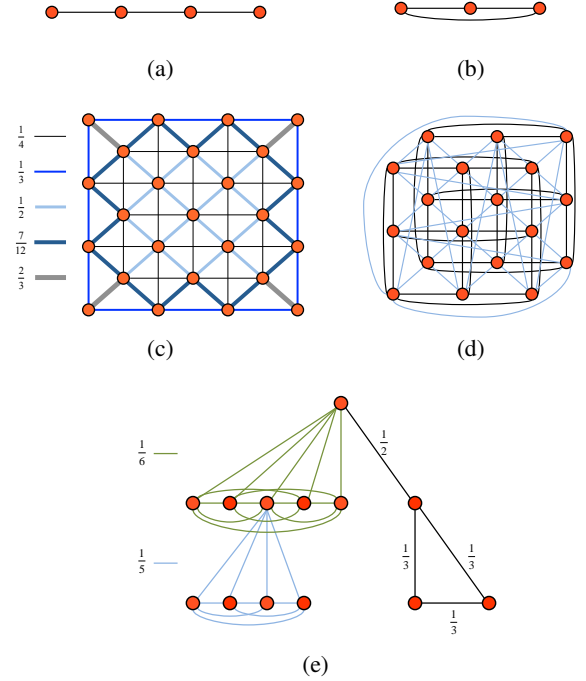


Fig. 3. Kron reductions of the graphs of Figure 1, with the blue vertices eliminated. If the weights of all of the edges in the original path and ring graphs are equal to 1, then the weights of all of the edges in the Kron-reduced graphs in (a) and (b) are equal to $\frac{1}{2}$. For the Kron reduction of the finite grid graph with the ends connected (d), the weights of the blue edges are equal to $\frac{1}{2}$ and the weights of the black edges are equal to $\frac{1}{4}$. The weights of Kron reductions of the finite grid graph with the ends unconnected (c) and the tree graph (e) are shown by color next to their respective Kron reductions.

C. Special Cases

As in Section III-B for the vertex selection process, we analyze the effect of Kron reduction on some special classes of graphs in order to (i) provide further intuition behind the graph reduction process; (ii) show that Kron reduction is a generalization of previously suggested techniques for reducing graphs belonging to these special classes; and (iii) highlight both the strengths and weaknesses of this choice of graph reduction.

1) *Paths*: The class of path graphs is closed under the sequential operations of largest eigenvector vertex selection and Kron reduction. That is, if we start with a path graph \mathcal{G}_P with N vertices, and select a subset \mathcal{V}_1 according to the polarity of the largest eigenvector of the graph Laplacian of \mathcal{G}_P , then the graph associated with $\mathcal{K}(\mathcal{G}_P, \mathcal{V}_1)$ is also a path graph. If the weights of the edges of the original graph are all equal to 1, the weights of the edges of the Kron-reduced graph are all equal to $\frac{1}{2}$. The Kron reduction of the graph shown in Figure 1(a) is shown in Figure 3(a).

2) *Rings*: The Kron reduction of the ring graph shown in Figure 1(b) is shown in Figure 3(b). It is a ring graph with half the number of vertices, and the weights of the edges are also halved in the Kron reduction process.

3) *Finite Grids*: The Kron reductions of the finite grid graphs shown in Figure 1(c,d) are shown in Figure 3(c,d). If we rotate Figure 3(c) clockwise by ninety degrees, we see that it is a subgraph of the 8-connected grid. We do not consider

graphs with infinite vertices here, but the combined effect of the largest eigenvector vertex selection and Kron reduction operations is to map an infinite (or very large) 4-connected grid to an 8-connected grid.

4) *Trees*: The Kron reduction of the tree shown in Figure 1(e) is shown in Figure 3(e). Note that the Kron reduction of a tree is not a tree because siblings in the original graph are connected in the Kron-reduced graph. However, it does have a particular structure, which is a union of complete subgraphs. For every node i in the reduced graph, and each one of that node's children j in the original graph, there is a complete subgraph in the reduced graph that includes i and all of the children of j (i.e., the grandchildren of i through j). If the weights of the edges in the original graph are equal to one, then the weights in each of the complete subgraphs are equal to one over the number of vertices in the complete subgraph.

This class of graphs highlights two of the main weaknesses of the Kron reduction: (i) it does not always preserve regular structural properties of the graph; and (ii) it does not always preserve the sparsity of the graph. We discuss a sparsity-enhancing modification in Section IV-D.

5) *k-Regular Bipartite Graphs*: In [18], Narang and Ortega consider connected and unweighted k -RBGs.⁷ They downsample by keeping one subset of the bipartition, and they construct a new graph on the downsampled vertices \mathcal{V}_1 by linking vertices in the reduced graph with an edge whose weight is equal to the number of their common neighbors in the original graph. If the vertices of the original graph are rearranged so that all the vertices in \mathcal{V}_1 have smaller indices than all the vertices in \mathcal{V}_1^c , the adjacency and Laplacian matrices of the original graph can be represented as:

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{W}_1 \\ \mathbf{W}_1^T & \mathbf{0} \end{bmatrix} \text{ and } \mathcal{L} = \begin{bmatrix} k\mathbf{I}_{\frac{N}{2}} & -\mathbf{W}_1 \\ -\mathbf{W}_1^T & k\mathbf{I}_{\frac{N}{2}} \end{bmatrix}. \quad (7)$$

Then for all $i, j \in \mathcal{V}_1$ (with $i \neq j$), the $(i, j)^{th}$ entry of the adjacency matrix of the reduced graph is given by

$$W_{ij}^{kRBG-reduced}(\mathcal{V}_1) = (W_1 W_1^T)_{ij}. \quad (8)$$

They also show that

$$\mathcal{L}^{kRBG-reduced}(\mathcal{V}_1) = k^2 \mathbf{I}_{\frac{N}{2}} - \mathbf{W}_1 \mathbf{W}_1^T. \quad (9)$$

Now we examine the Kron reduction of k -RBGs. The Kron-reduced Laplacian is given by:

$$\begin{aligned} \mathcal{L}^{Kron-reduced}(\mathcal{V}_1) &= \mathcal{L}_{\mathcal{V}_1, \mathcal{V}_1} - \mathcal{L}_{\mathcal{V}_1, \mathcal{V}_1^c} \mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1}^{-1} \mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1} \\ &= k\mathbf{I}_{\frac{N}{2}} - (-\mathbf{W}_1)(k\mathbf{I}_{\frac{N}{2}})^{-1}(-\mathbf{W}_1^T) \\ &= k\mathbf{I}_{\frac{N}{2}} - \frac{1}{k} \mathbf{W}_1 \mathbf{W}_1^T, \end{aligned}$$

which is a constant factor $\frac{1}{k}$ times the reduced Laplacian (9) of [18]. So, up to a constant factor, the Kron reduction is a generalization of the graph reduction method presented in [18] for the special case of regular bipartite graphs.

⁷Although [18] considers unweighted k -RBGs, the following statements also apply to weighted k -RBGs if we extend the definition of the reduced adjacency matrix (8) to weighted graphs.

Algorithm 1 Spectral Sparsification [30]

Inputs: $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}, Q$

Output: \mathbf{W}'

- 1: Initialize $\mathbf{W}' = \mathbf{0}$
- 2: **for** $q = 1, 2, \dots, Q$ **do**
- 3: Choose a random edge $e = (i, j)$ of \mathcal{E} according to the probability distribution

$$p_e = \frac{d_{R\mathcal{G}}(i, j)W_{ij}}{\sum_{e=(m, n) \in \mathcal{E}} d_{R\mathcal{G}}(m, n)W_{mn}}$$

- 4: $\mathbf{W}'_{ij} = \mathbf{W}'_{ij} + \frac{W_{ij}}{Qp_e}$
 - 5: **end for**
-

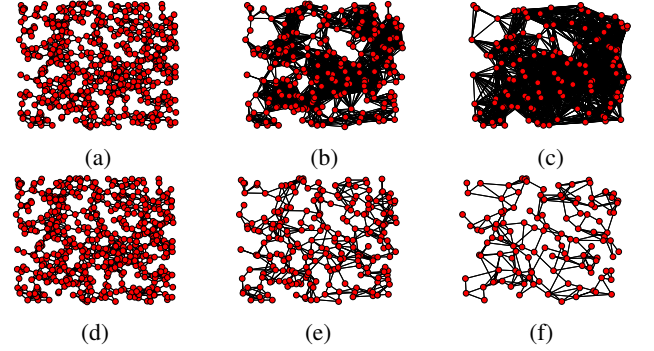


Fig. 4. Incorporation of a spectral sparsification step into the graph reduction. (a)-(c) Repeated largest eigenvector downsampling and Kron reduction of a sensor network graph. (d)-(f) The same process with the spectral sparsification of [30] used immediately after each Kron reduction.

D. Graph Sparsification

As a consequence of property (K5), repeated Kron reduction often leads to denser and denser graphs. We have already seen this loss of sparsity in Section IV-C4, and this phenomenon is even more evident in larger, less regular graphs. In addition to computational drawbacks, the loss of sparsity can be important, because if the reduced graphs become too dense, they may not effectively capture the local connectivity information that is important for processing signals on the graph. Therefore, in many situations, it is advantageous to perform graph sparsification immediately after the Kron reduction as part of the overall graph reduction phase.

There are numerous ways to perform graph sparsification. In this paper, we use a straightforward spectral sparsification algorithm of Spielman and Strivastava [30], which is described in Algorithm 1. This sparsification method pairs nicely with the Kron reduction, because [30] shows that for large graphs and an appropriate choice of the number of samples Q , the graph Laplacian spectrum and resistance distances between vertices are approximately preserved with high probability. In Figure 4, we show an example of repeated downsampling followed by Kron reduction and spectral sparsification.

E. Alternative Graph Reduction Methods

First, we mention some alternative graph reduction methods:

- 1) In [8], Narang and Ortega define a reduced graph via the weighted adjacency matrix by taking $\mathbf{W}^{(j+1)} = ([\mathbf{W}^{(j)}]^2)_{\mathcal{V}_1, \mathcal{V}_1}$. Here, $[\mathbf{W}^{(j)}]^2$ represents the 2-hop adjacency matrix of the original graph. The reduced Laplacian can then be defined as $\mathcal{L}^{(j+1)} = \mathbf{D}^{(j+1)} - \mathbf{W}^{(j+1)}$, where $\mathbf{D}^{(j+1)}$ is computed from $\mathbf{W}^{(j+1)}$. However, there are a number of undesirable properties of this reduction method. First, and perhaps foremost, the reduction method does not always preserve connectivity. Second, self-loops are introduced at every vertex in the reduced graph. Third, vertices in the selected subset that are connected by an edge in the original graph may not share an edge in the reduced graph. Fourth, the spectrum of the reduced graph Laplacian is not necessarily contained in the spectrum of the original graph Laplacian.
- 2) Ron et al. [13] assign a fraction P_{ij} of each vertex i in the original graph to each vertex $j \in \mathcal{V}_1$ in the reduced graph that is close to i in the original graph (in terms of an algebraic distance). The assignment satisfies $\sum_{j \in \mathcal{V}_1} P_{ij} = 1$ for all $i \in \mathcal{V}$ and $P_{jj} = 1$ for all $j \in \mathcal{V}_1$, and for each eliminated vertex $i \in \mathcal{V}_1^c$, an upper limit is placed on the number of vertices $j \in \mathcal{V}_1$ that have $P_{ij} > 0$. Then for all $j, j' \in \mathcal{V}_1$,

$$W_{jj'}^{\text{reduced}} := \sum_{m, n \in \mathcal{V}, m \neq n} P_{mj} P_{nj'} W_{mn}.$$

Next, we mention some graph coarsening (also called coarse-graining) methods that combine graph downsampling and reduction into a single operation by forming aggregate nodes at each resolution, rather than keeping a strict subset of original vertices. The basic approach of these methods is to partition the original set of vertices into clusters, represent each cluster of vertices in the original graph with a single vertex in the reduced graph, and then use the original graph to form edges and weights that connect the representative vertices in the reduced graph. Some examples include:

- 1) Lafon and Lee [31] cluster based on diffusion distances and form new edge weights based on random walk transition probabilities.
- 2) The Lean Algebraic Multigrid (LAMG) method of [32] builds each coarser Laplacian by selecting seed nodes, assigning each non-seed node to be aggregated with exactly one seed node, and setting the weights between two new seed nodes j and j' to be

$$W_{jj'}^{\text{reduced}} := \sum_{i \in \mathcal{V}_j} \sum_{i' \in \mathcal{V}_{j'}} W_{ii'},$$

where \mathcal{V}_j and $\mathcal{V}_{j'}$ are the sets of vertices that have been aggregated with seeds j and j' , respectively. The seed assignments are based on an affinity measure that approximates short time diffusion distances, rather than the algebraic distances used in [13].

- 3) Multilevel clustering algorithms such as those presented in [33], [34] often use greedy coarsening algorithms such as heavy edge matching or max-cut coarsening (see [33, Section 3] and [34, Section 5.1] for details).

For more thorough reviews of the graph partitioning and coarsening literature, see, e.g., [11], [13], [35].

To summarize, given a graph \mathcal{G} and a desired number of resolution levels, in order to generate a multiresolution of graphs, we only need to choose a downsampling operator and a graph reduction method. In the remainder of the paper, we use the largest eigenvector downsampling operator and the Kron reduction with the extra sparsification step to generate graph multiresolutions such as the three level one shown in Figure 4(d)-(f). Note that graph multiresolutions generated in this fashion are completely independent of any signals residing on the graph.

V. FILTERING AND INTERPOLATION OF GRAPH SIGNALS

Equipped with a graph multiresolution, we now proceed to analyze signals residing on the finest graph in the multiresolution. The two key graph signal processing components we use in the proposed transform are a generalized filtering operator and an interpolation operator for signals on graphs. Filters are commonly used in classical signal processing analysis to separate a signal into different frequency bands. In this section, we review how to extend the notion of filtering to graph signals. We focus on graph spectral filtering, which leverages the eigenvalues and eigenvectors of Laplacian operators from spectral graph theory [36] to capture the geometric structure of the underlying graph data domain. We then discuss different methods to interpolate from a signal residing on a coarser graph to a signal residing on a finer graph whose vertices are a superset of those in the coarser graph.

A. Graph Spectral Filtering

A *graph signal* is a function $f : \mathcal{V} \rightarrow \mathbb{R}$ that associates a real value to each vertex of the graph. Equivalently, we can view a graph signal as a vector $f \in \mathbb{R}^N$.

In frequency filtering, we represent signals as linear combinations of a set of signals and amplify or attenuate the contributions of different components. In classical signal processing, the set of component signals are usually the complex exponentials, which carry a notion of frequency and give rise to the Fourier transform. In graph signal processing, it is most common to choose the graph Fourier expansion basis to be the eigenvectors of the combinatorial or normalized graph Laplacian operators. This is because the spectra of these graph Laplacians also carry a notion of frequency (see, e.g., [2, Figure 3]), and their eigenvectors are the graph analogs to the complex exponentials, which are the eigenfunctions of the classical Laplacian operator.

More precisely, the graph Fourier transform with the combinatorial graph Laplacian eigenvectors as a basis is

$$\hat{f}(\lambda_\ell) := \langle \mathbf{f}, \mathbf{u}_\ell \rangle = \sum_{i=1}^N f(i) u_\ell^*(i), \quad (10)$$

and a *graph spectral filter*, which we also refer to as a kernel, is a real-valued mapping $\hat{h}(\cdot)$ on the spectrum of graph Laplacian eigenvalues. Just as in classical signal processing, the effect of the filter is multiplication in the Fourier domain:

$$\hat{f}_{\text{out}}(\lambda_\ell) = \hat{f}_{\text{in}}(\lambda_\ell) \hat{h}(\lambda_\ell), \quad (11)$$

or, equivalently, taking an inverse graph Fourier transform,

$$f_{out}(i) = \sum_{\ell=0}^{N-1} \hat{f}_{in}(\lambda_\ell) \hat{h}(\lambda_\ell) u_\ell(i). \quad (12)$$

We can also write the filter in matrix form as $\mathbf{f}_{out} = \mathbf{H}\mathbf{f}_{in}$, where \mathbf{H} is a matrix function [37]

$$\mathbf{H} = \hat{h}(\mathcal{L}) = \mathbf{U}[\hat{h}(\mathbf{\Lambda})]\mathbf{U}^*, \quad (13)$$

where $\hat{h}(\mathbf{\Lambda})$ is a diagonal matrix with the elements of the diagonal equal to $\{\hat{h}(\lambda_\ell)\}_{\ell=0,1,\dots,N-1}$. We can also use the normalized graph Laplacian eigenvectors as the graph Fourier basis, and simply replace \mathcal{L} , λ_ℓ , and \mathbf{u}_ℓ by $\tilde{\mathcal{L}}$, λ_ℓ , and $\tilde{\mathbf{u}}_\ell$ in (10)-(13). A discussion of the benefits and drawbacks of each of these choices for the graph Fourier basis is included in [2].

B. Alternative Filtering Methods for Graph Signals

We briefly mention two alternative graph filtering methods:

- 1) We can filter a graph signal directly in the vertex domain by writing the output at a given vertex i as a linear combination of the input signal components in a neighborhood of i . Graph spectral filtering with an order K polynomial kernel can be viewed as filtering in the vertex domain with the component of the output at vertex i written as a linear combination of the input signal components in a K -hop neighborhood of i (see [2] for more details).
- 2) Other choices of filtering bases can be used in place of \mathcal{L} in (13). For example, in [38], [39], Sandryhaila and Moura examine filters that are polynomial functions of the adjacency matrix.

C. Interpolation on Graphs

Given the values of a graph signal on a subset, \mathcal{V}_1 , of the vertices, the interpolation problem is to infer the values of the signal on the remaining vertices, $\mathcal{V}_1^c = \mathcal{V} \setminus \mathcal{V}_1$. It is usually assumed that the signal to be interpolated is smooth. This assumption is always valid in our construction, because the signals we wish to interpolate have been smoothed via a lowpass graph spectral filter.

We use the interpolation model of Pesenson [40], which is to interpolate the missing values, $\mathbf{f}_{\mathcal{V}_1^c}$, of the signal \mathbf{f} by writing the interpolant as a linear combination of the Green's functions of the regularized graph Laplacian, $\tilde{\mathcal{L}} := \mathcal{L} + \epsilon \mathbf{I}$, that are centered at the vertices in \mathcal{V}_1 .⁸ That is, the interpolation is given by

$$\mathbf{f}_{\text{interp}} = \sum_{j \in \mathcal{V}_1} \alpha[j] \varphi_j = \Phi_{\mathcal{V}_1} \alpha, \quad (14)$$

where φ_j is a Green's kernel $\hat{g}(\lambda_\ell) = \frac{1}{\lambda_\ell + \epsilon}$ translated to center vertex j (see [5], [2] for more on generalized translation on graphs):

$$\varphi_j = T_j g = \hat{g}(\mathcal{L}) \delta_j = \mathbf{U}[\hat{g}(\mathbf{\Lambda})]\mathbf{U}^* \delta_j = \sum_{\ell=0}^{|\mathcal{V}|-1} \frac{1}{\lambda_\ell + \epsilon} u_\ell^*(j) \mathbf{u}_\ell,$$

⁸Pesenson's model and subsequent analysis in [40] is for any power $\tilde{\mathcal{L}}^t$ of the regularized Laplacian, but we stick to $t = 1$ throughout for simplicity.

and $\Phi_{\mathcal{V}_1}$ is a $|\mathcal{V}| \times |\mathcal{V}_1|$ matrix with columns $\{\varphi_j\}_{j \in \mathcal{V}_1}$. We display three such translated Green's functions in Figure 5. Note that these are localized functions satisfying the property

$$\tilde{\mathcal{L}} \varphi_j = \left(\mathbf{U}[\mathbf{\Lambda} + \epsilon \mathbf{I}] \mathbf{U}^* \right) \left(\mathbf{U}[\hat{g}(\mathbf{\Lambda})] \mathbf{U}^* \delta_j \right) = \tilde{\mathcal{L}} \tilde{\mathcal{L}}^{-1} \delta_j = \delta_j.$$

The parameter ϵ controls the spread of the translated Green's functions around their center vertices, with a smaller value of ϵ leading to larger spreads.

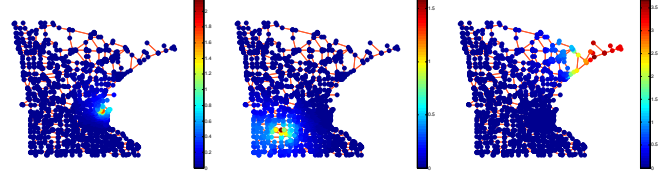


Fig. 5. Three localized Green's functions, centered at different vertices on the Minnesota graph.

To perform the interpolation, we find the coefficients α such that the linear combination of Green's functions restricted to the subset \mathcal{V}_1 of vertices yields the signal values on this subset:

$$\mathbf{f}_{\mathcal{V}_1} = \Phi_{\mathcal{V}_1, \mathcal{V}_1} \alpha_*, \quad (15)$$

where $\mathbf{f}_{\mathcal{V}_1}$ is a vector containing the values of the signal on the vertices in \mathcal{V}_1 , and $\Phi_{\mathcal{V}_1, \mathcal{V}_1}$ is the submatrix that contains only the rows of $\Phi_{\mathcal{V}_1}$ corresponding to the vertices in \mathcal{V}_1 . To compute $\alpha_* = \Phi_{\mathcal{V}_1, \mathcal{V}_1}^{-1} \mathbf{f}_{\mathcal{V}_1}$, we note that

$$\begin{aligned} \Phi_{\mathcal{V}_1, \mathcal{V}_1} &= (\tilde{\mathcal{L}}^{-1})_{\mathcal{V}_1, \mathcal{V}_1} \\ &= \left[\tilde{\mathcal{L}}_{\mathcal{V}_1, \mathcal{V}_1} - \tilde{\mathcal{L}}_{\mathcal{V}_1, \mathcal{V}_1^c} (\tilde{\mathcal{L}}_{\mathcal{V}_1^c, \mathcal{V}_1^c})^{-1} \tilde{\mathcal{L}}_{\mathcal{V}_1^c, \mathcal{V}_1} \right]^{-1}, \end{aligned}$$

where the last equality follows from the Matrix Inversion Lemma. Thus, the cost to compute

$$\alpha_* = \left[\tilde{\mathcal{L}}_{\mathcal{V}_1, \mathcal{V}_1} - \tilde{\mathcal{L}}_{\mathcal{V}_1, \mathcal{V}_1^c} (\tilde{\mathcal{L}}_{\mathcal{V}_1^c, \mathcal{V}_1^c})^{-1} \tilde{\mathcal{L}}_{\mathcal{V}_1^c, \mathcal{V}_1} \right]^{-1} \mathbf{f}_{\mathcal{V}_1} \quad (16)$$

is dominated by the solution of a sparse diagonally dominant set of $|\mathcal{V}| - |\mathcal{V}_1|$ linear equations. Finally, we can substitute α_* back into (14) to find the interpolated signal, whose values on \mathcal{V}_1 of course correspond to the given values $\mathbf{f}_{\mathcal{V}_1}$. This last calculation can be approximated efficiently by upsampling α_* to \mathcal{V} , and filtering it with the Green's kernel $\hat{g}(\cdot)$ using the polynomial approximation algorithm discussed in Section VII-C. Pesenson [40] refers to the interpolated signal $\mathbf{f}_{\text{interp}}$ as a *variational spline*.

As a brief aside, we note that the matrix in (16) is exactly the Kron reduction $\mathcal{K}(\tilde{\mathcal{L}}, \mathcal{V}_1)$ of the regularized graph Laplacian on the set of vertices \mathcal{V}_1 where the signal values are known.

As with downsampling and graph reduction, there are a number of different methods to interpolate smooth signals on graphs, and we briefly mention a few alternatives here:

- 1) Graph-based semi-supervised learning methods, such as Tikhonov regularization (see, e.g., [41]-[46]). Like many of those methods, the spline interpolation method we use can be viewed as solving an optimization problem (see [40]). One main difference is that in our multiscale transform, we will usually have approximately half of the signal values available, whereas the number of

signal values available in the semi-supervised learning literature is usually assumed to be significantly smaller.

- 2) Grady and Schwartz's anisotropic interpolation solves the system of equations $\mathcal{L}_{V_1^c, V_1^c} \mathbf{f}_{V_1^c} = -\mathcal{L}_{V_1^c, V_1} \mathbf{f}_{V_1}$ in order to minimize $\mathbf{f}_{\text{interp}}^T \mathcal{L} \mathbf{f}_{\text{interp}}$ subject to $f_{\text{interp}}(i) = f(i)$ for all $i \in \mathcal{V}$.
- 3) Rather than using the Green's functions as the interpolating functions in (14), Narang et al. [47], [48] use the eigenvectors of the normalized graph Laplacian associated with the lowest eigenvalues.

VI. A PYRAMID TRANSFORM FOR SIGNALS ON GRAPHS

We are finally ready to combine the downsampling, graph reduction, filtering, and interpolation operations from the previous three sections to define a multiscale pyramid transform for signals on graphs. After reviewing the classical Laplacian pyramid, we summarize our extension, first for a single level and then for the whole pyramid.

A. The Classical Laplacian Pyramid

In [15], Burt and Adelson introduce the Laplacian pyramid. Originally designed with image coding in mind, the Laplacian pyramid is a multiresolution transform that is applicable to any regularly-sampled signal in time or space. A schematic representation of a single level of the Laplacian pyramid is shown in Figure 6(a). At each level of the pyramid, an input signal \mathbf{x} is lowpass filtered (\mathbf{H}) and then downsampled to form a coarse approximation of the original signal. A prediction of the input signal is then formed by upsampling and lowpass filtering (\mathbf{G}) the coarse approximation. The prediction error \mathbf{y} , i.e., the difference between the input signal and the prediction based on the coarse approximation, is stored for reconstruction. This process is iterated, with the coarse approximation that is output at the previous level acting as the input to the next level. The sequence $\{\mathbf{x}^{(i)}\}_{i=1,2,\dots}$ represents a series of coarse approximations of the original signal $\mathbf{x}^{(0)}$ at decreasing resolutions. If, for example, the downsampling is by a factor of 2 and the original signal $\mathbf{x}^{(0)} \in \mathbb{R}^N$, then $\mathbf{x}^{(i)} \in \mathbb{R}^{2^{-i}N}$.

The lowest resolution approximation, $\mathbf{x}^{(J)}$, of the original signal and the sequence of prediction errors, $\{\mathbf{y}^{(i)}\}_{i=0,1,\dots,J-1}$ are stored and/or transmitted for reconstruction. Therefore, the Laplacian pyramid is an overcomplete transform, as a J -level pyramid downsampled by a factor of κ at each level maps an input of dimension N into $N \left(\frac{1 - \frac{1}{\kappa}}{1 - \frac{1}{\kappa^{J+1}}} \right)$ transform coefficients. On the other hand, the Laplacian pyramid has a number of desirable properties. First, perfect reconstruction of the original signal is possible for any choice of filtering operations \mathbf{H} and \mathbf{G} . It is easy to see that upsampling $\mathbf{x}^{(j+1)}$, filtering it by \mathbf{G} , and summing with the prediction error $\mathbf{y}^{(j)}$ yields the finer approximation $\mathbf{x}^{(j)}$. Second, the prediction errors usually have less entropy than the original signal of the same dimension, enabling further compression. Third, it is computationally efficient to implement the transform, due to the hierarchical and local nature of the computations.

For some applications, the reconstruction process starts with noisy transform coefficients, due to communication noise,

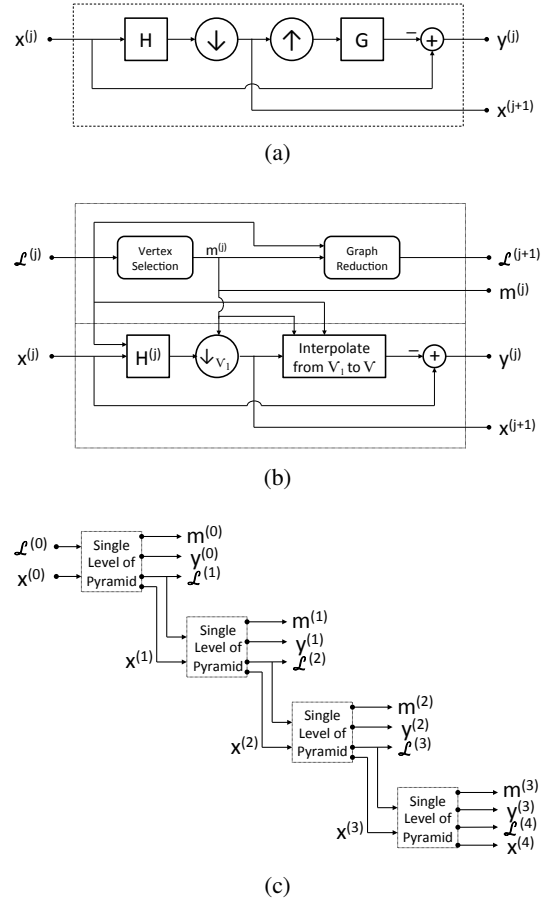


Fig. 6. (a) A single level of the classical Laplacian pyramid of Burt and Adelson [15]. (b) A single level of the pyramid scheme proposed for signals defined on graphs. The top half consisting of the graph downsampling and reduction is specific to the graph (but the same for every signal on that graph), while the bottom half is specific to each graph signal. (c) A multilevel pyramid for signals defined on graphs.

quantization error, or other types of processing. In this case, Do and Vetterli [49], [50] show that Burt and Adelson's single-level synthesis operator is not optimal in terms of the reconstruction error. The single-level analysis operator that transforms a coarse approximation $\mathbf{x}^{(j)}$ into the next coarse approximation $\mathbf{x}^{(j+1)}$ and prediction error $\mathbf{y}^{(j)}$ is a *frame operator* (see, e.g., [51]). Frame operators have an infinite number of left inverses, but the one that minimizes the reconstruction error is the pseudoinverse of the analysis operator at each level of the pyramid. In Section VI-D, we also use the pseudoinverse of the analysis operator as the synthesis operator at each level of our Laplacian pyramid for signals on graphs. Thus, we defer a more precise mathematical treatment of the optimal reconstruction until then.

B. The Analysis Operator for a Single Level of the Pyramid for Signals on Graphs

We now extend the classical Laplacian pyramid to a pyramid transform for signals on graphs. At each level of the pyramid, the inputs are the current coarse approximation, $\mathbf{x}^{(j)} \in \mathbb{R}^{N^{(j)}}$, of the original signal, and the current graph Laplacian $\mathcal{L}^{(j)} \in \mathbb{R}^{N^{(j)}} \times \mathbb{R}^{N^{(j)}}$. We use the polarity of the largest eigenvector

of $\mathcal{L}^{(j)}$, as described in Section III, to select the vertices of the graph on which to form the reduced graph at resolution scale $j+1$. We denote the output of the vertex selection process by $\mathbf{m}^{(j)} \in \mathbb{R}^{N^{(j)}}$, where

$$m^{(j)}(i) = \begin{cases} 1, & \text{if vertex } i \text{ is selected to be included in the reduced graph} \\ 0, & \text{otherwise} \end{cases}$$

The vertex selection vector $\mathbf{m}^{(j)}$ is used both to reduce the graph and to define the downsampling and interpolation operators. For graph reduction, we use Kron reduction followed by spectral sparsification, as discussed in Section IV. The graph Laplacian of the reduced graph is given by

$$\begin{aligned} \mathcal{L}^{(j+1)} &= \mathcal{S} \left(\mathcal{K} \left(\mathcal{L}^{(j)}, \mathbf{m}^{(j)} \right) \right) \\ &= \mathcal{S} \left(\mathcal{L}_{\mathcal{V}_1^{(j)}, \mathcal{V}_1^{(j)}}^{(j)} - \mathcal{L}_{\mathcal{V}_1^{(j)}, \mathcal{V}_1^{(j)c}}^{(j)} \left[\mathcal{L}_{\mathcal{V}_1^{(j)c}, \mathcal{V}_1^{(j)c}}^{(j)} \right]^{-1} \mathcal{L}_{\mathcal{V}_1^{(j)c}, \mathcal{V}_1^{(j)}}^{(j)} \right), \end{aligned}$$

where $\mathcal{V}_1^{(j)} := \{i \in \{1, 2, \dots, N^{(j)}\} : m^{(j)}(i) = 1\}$ and \mathcal{S} is the spectral sparsification operator. The reduced graph $\mathcal{L}^{(j+1)}$ and the vertex selection vector $\mathbf{m}^{(j)}$, both of which need to be stored for reconstruction, are the first two outputs of the transform at each level of the pyramid.

The other two outputs are the coarse approximation vector $\mathbf{x}^{(j+1)}$ at the coarser resolution scale $j+1$, and the next prediction error vector $\mathbf{y}^{(j)}$. The coarse approximation vector is computed in a similar fashion as in the classical Laplacian pyramid, with the filtering and downsampling operators replaced by their graph analogs. The downsampling operator \downarrow corresponds to matrix multiplication by the matrix $\mathbf{S}_d^{(j)} \in \mathbb{R}^{N^{(j+1)}} \times \mathbb{R}^{N^{(j)}}$, where $\mathbf{S}_d^{(j)} := [\mathbf{I}_{N^{(j)}}]_{\mathcal{V}_1^{(j)}, \mathcal{V}_1^{(j)}}$ and $\mathbf{I}_{N^{(j)}}$ is an $N^{(j)} \times N^{(j)}$ identity matrix. For example, if $\mathbf{m}^{(j)} = (1, 1, 0, 0, 1, 0)^T$, then

$$\mathbf{S}_d^{(j)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Rather than upsample and filter with another lowpass filter G as in the classical pyramid, we compute the prediction error with the spline interpolation method of Section V-C. While this adds some additional computational cost, our experiments showed that this change significantly improves the predictions, and in turn the sparsity of the transform coefficients. So the single-level analysis operator $\mathbf{T}_a^{(j)} : \mathbb{R}^{N^{(j)}} \rightarrow \mathbb{R}^{N^{(j)}+N^{(j+1)}}$ is given by

$$\begin{aligned} \mathbf{T}_a^{(j)} \mathbf{x}^{(j)} &:= \begin{bmatrix} \mathbf{S}_d^{(j)} \mathbf{H}^{(j)} \\ \mathbf{I}_{N^{(j)}} - \Phi_{\mathcal{V}_1}^{(j)} \left(\Phi_{\mathcal{V}_1, \mathcal{V}_1}^{(j)} \right)^{-1} \mathbf{S}_d^{(j)} \mathbf{H}^{(j)} \end{bmatrix} \mathbf{x}^{(j)} \\ &= \begin{bmatrix} \mathbf{x}^{(j+1)} \\ \mathbf{y}^{(j)} \end{bmatrix} \end{aligned} \quad (17)$$

In (17), the filtering operator is $\mathbf{H}^{(j)} = \mathbf{U}^{(j)} [\hat{h}(\boldsymbol{\Lambda}^{(j)})] \mathbf{U}^{(j)*}$.

A schematic representation of a single level of the pyramid for signals on graphs is shown in Figure 6(b). Note that while we use the polarity of the largest eigenvector of the graph Laplacian as a vertex selection method, the sparsified Kron reduction as a graph reduction method, the eigenvectors

of the graph Laplacian as a filtering basis, and the spline interpolation method with translated Green's functions, other vertex selection, graph reduction, filtering, and interpolation methods can be substituted into these blocks without affecting the perfect reconstruction property of the scheme.

C. The Multilevel Pyramid

The multilevel pyramid for signals on graphs is shown in Figure 6(c). As in the classical case, we iterate the single-level analysis on the downsampled output of the lowpass channel. In (17), the filtering operators $\mathbf{H}^{(j)}$ necessarily depend on the resolution scale j through their dependence on the eigenvectors and eigenvalues of $\mathcal{L}^{(j)}$. However, the filter $\hat{h} : [0, \lambda_{\max}^{(0)}] \rightarrow \mathbb{R}$ may be fixed across all levels of the pyramid due to property (6) of the Kron reduction, which guarantees that the spectrum of $\mathcal{L}^{(j)}$ is contained in $[0, \lambda_{\max}^{(0)}]$ for all j .⁹

D. Synthesis Operators for a Single Level and Optimal Reconstruction

Just as is the case in the classical Laplacian pyramid, for any choices of filter $\hat{h}(\cdot)$, we can reconstruct $\mathbf{x}^{(j)}$ perfectly from $\mathbf{x}^{(j+1)}$ and $\mathbf{y}^{(j)}$ with a single-level synthesis operator, $\mathbf{T}_s^{(j)} : \mathbb{R}^{N^{(j)}+N^{(j+1)}} \rightarrow \mathbb{R}^{N^{(j)}}$, given by

$$\mathbf{T}_s^{(j)} \begin{bmatrix} \mathbf{x}^{(j+1)} \\ \mathbf{y}^{(j)} \end{bmatrix} := \begin{bmatrix} \Phi_{\mathcal{V}_1}^{(j)} \left(\Phi_{\mathcal{V}_1, \mathcal{V}_1}^{(j)} \right)^{-1} & \mathbf{I}_{N^{(j)}} \end{bmatrix} \begin{bmatrix} \mathbf{x}^{(j+1)} \\ \mathbf{y}^{(j)} \end{bmatrix}. \quad (18)$$

Note that in order to apply the synthesis operator $\mathbf{T}_s^{(j)}$, we need access to both $\mathcal{L}^{(j)}$ and $\mathbf{m}^{(j)}$, which are used for the interpolation.

It is straightforward to check that $\mathbf{T}_s^{(j)} \mathbf{T}_a^{(j)} = \mathbf{I}_{N^{(j)}}$, guaranteeing perfect reconstruction. However, once again analogously to the classical Laplacian pyramid, when noise is introduced into the stored coefficients $\mathbf{x}^{(j+1)}$ and $\mathbf{y}^{(j)}$, we would like to reconstruct with the pseudoinverse of $\mathbf{T}_a^{(j)}$:

$$\mathbf{T}_a^{(j)\dagger} := \left(\mathbf{T}_a^{(j)*} \mathbf{T}_a^{(j)} \right)^{-1} \mathbf{T}_a^{(j)*}. \quad (19)$$

If $\tilde{\mathbf{x}}^{(j+1)}$ and $\tilde{\mathbf{y}}^{(j)}$ are the noisy versions of $\mathbf{x}^{(j+1)}$ and $\mathbf{y}^{(j)}$, respectively, then $\tilde{\mathbf{x}}^{(j)} = \mathbf{T}_a^{(j)\dagger} \begin{bmatrix} \tilde{\mathbf{x}}^{(j+1)} \\ \tilde{\mathbf{y}}^{(j)} \end{bmatrix}$ minimizes the reconstruction error. That is,

$$\mathbf{T}_a^{(j)\dagger} \begin{bmatrix} \tilde{\mathbf{x}}^{(j+1)} \\ \tilde{\mathbf{y}}^{(j)} \end{bmatrix} = \underset{\mathbf{x} \in \mathbb{R}^{N^{(j)}}}{\operatorname{argmin}} \left\| \mathbf{T}_a^{(j)} \mathbf{x} - \begin{bmatrix} \tilde{\mathbf{x}}^{(j+1)} \\ \tilde{\mathbf{y}}^{(j)} \end{bmatrix} \right\|_2. \quad (20)$$

For large graphs, rather than form and compute the matrix inverse in (19), it may be more computationally efficient to approximate the left-hand side of (20) with Landweber iteration [54, Theorem 6.1].

To summarize, for a multilevel pyramid for signals on graphs that has J resolution levels, the reconstruction process

⁹If we include spectral sparsification in the graph reduction step, then the spectrum of $\mathcal{L}^{(j)}$ is not guaranteed to be contained in $[0, \lambda_{\max}^{(0)}]$; however, for large graphs, the results of [30] limit the extent to which the maximum graph Laplacian eigenvalue can increase beyond $\lambda_{\max}^{(0)}$. In any case, we usually define the filters on the entire positive real line, in which case there is no issue with the length of the spectrum increasing at successive levels of the pyramid.

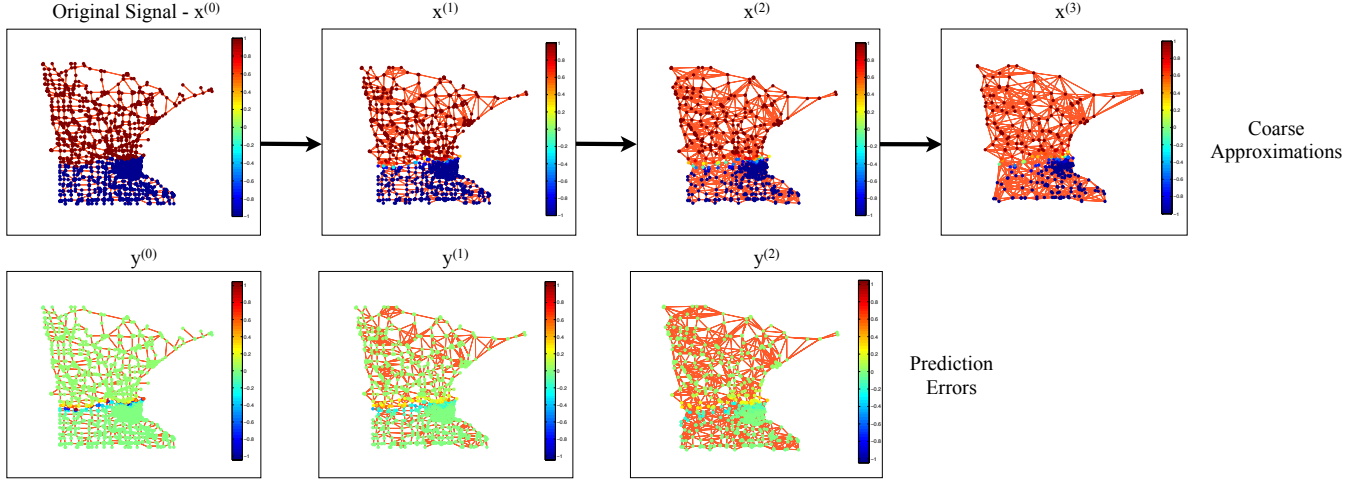


Fig. 7. Three-level pyramid analysis of a piecewise-constant signal on the Minnesota road network of [52]. The number of vertices is reduced in successive coarse approximations from $N = 2642$ to 1334 to 669 to 337. The overall redundancy factor of the transform in this case is 1.89.

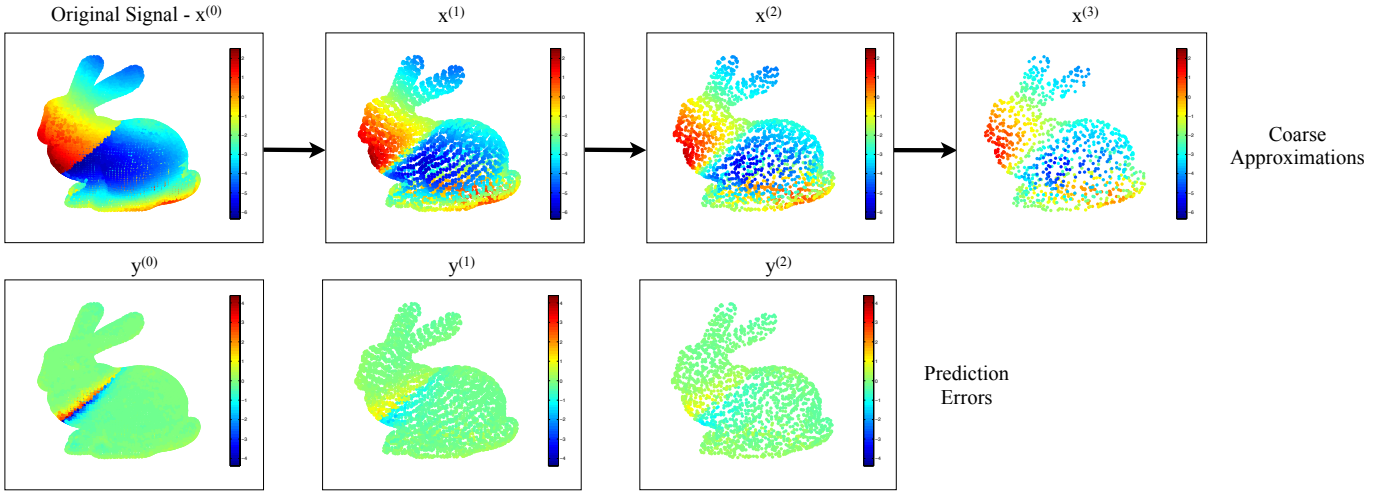


Fig. 8. Three-level pyramid analysis of a piecewise-smooth signal on the Stanford Bunny graph of [53]. The number of vertices is reduced in successive coarse approximations from $N = 8170$ to 4108 to 2050 to 1018. The overall redundancy factor of the transform in this case is 1.88.

begins with knowledge of $\mathbf{m}^{(0)}, \mathbf{m}^{(1)}, \dots, \mathbf{m}^{(J-1)}, \mathcal{L}^{(0)}, \mathcal{L}^{(1)}, \dots, \mathcal{L}^{(J-1)}, \mathbf{y}^{(0)}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(J-1)}$, and $\mathbf{x}^{(J)}$. We first use $\mathbf{m}^{(J-1)}$ and $\mathcal{L}^{(J-1)}$ to apply $\mathbf{T}_a^{(J-1)\dagger}$ to $\begin{bmatrix} \mathbf{x}^{(J)} \\ \mathbf{y}^{(J-1)} \end{bmatrix}$. The result is $\tilde{\mathbf{x}}^{(J-1)}$. We then use $\mathbf{m}^{(J-2)}$ and $\mathcal{L}^{(J-2)}$ to apply $\mathbf{T}_a^{(J-2)\dagger}$ to $\begin{bmatrix} \tilde{\mathbf{x}}^{(J-1)} \\ \mathbf{y}^{(J-2)} \end{bmatrix}$ in order to compute $\tilde{\mathbf{x}}^{(J-2)}$. We iterate this process J times until we finally compute $\tilde{\mathbf{x}}^{(0)}$, which, if no noise has been introduced into the transform coefficients, is equal to the original signal $\mathbf{x}^{(0)}$.

E. Illustrative Examples

In Figure 7, we use the proposed pyramid transform to analyze a piecewise-constant signal (the Fiedler vector \mathbf{u}_1) on the Minnesota road network of [52]. After each Kron reduction at stage j , we use the spectral sparsification of Algorithm 1 with Q set to the integer closest to $4N^{(j)} \log(N^{(j)})$. The graph spectral filter is $\hat{h}(\lambda_\ell) := \frac{0.5}{0.5 + \lambda_\ell}$, and we take $\epsilon = 0.005$ to form the regularized Laplacian for interpolation. We see

that the prediction error coefficients are extremely sparse, and those with non-zero magnitudes are concentrated around the discontinuity in the graph signal.

In Figure 8, we apply the transform to a piecewise smooth signal on the Stanford Bunny graph [53] with 8170 vertices. The signal is comprised of two different polynomials of the coordinates: $\bar{x}^2 + \bar{y} - 2\bar{z}$ on the head and ears of the bunny, and $\bar{x} - \bar{y} + 3\bar{z}^2 - 5$ on the lower body, where $(\bar{x}, \bar{y}, \bar{z})$ are the physical coordinates of each vertex. Since the graph is larger and the graph Laplacian spectrum is wider than the previous example, we take Q to be the integer closest to $16N^{(j)} \log(N^{(j)})$, the lowpass filter to be $\hat{h}(\lambda_\ell) := \frac{5}{5 + \lambda_\ell}$, and $\epsilon = 0.05$ for interpolation.

As a compression example, we hard threshold the 15346 pyramid transform coefficients of the signal, keeping only the 2724 with the largest magnitudes (approximately 1/3 the size of the original signal) and setting the rest to 0. Reconstruction of the signal from the thresholded coefficients yields an error

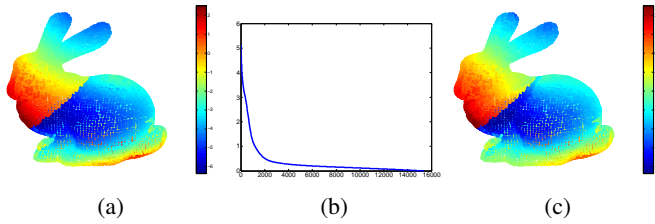


Fig. 9. Compression example. (a) The original piecewise-smooth signal with a discontinuity on the Stanford bunny [53]. (b) The sorted magnitudes of the 15346 pyramid transform coefficients. (c) The reconstruction from the 2724 coefficients with the largest magnitudes, using the least squares synthesis (19).

$\frac{\|f_{\text{reconstruction}} - f\|_2}{\|f\|_2}$ of 0.145 with the direct synthesis (18), and an error of 0.086 with the least squares synthesis (19). Figure 9 shows the original signal, the sorted magnitudes of the transform coefficients, and the reconstruction with the least squares synthesis operator.

F. Comparison with Other Transforms

A thorough comparison of our proposed pyramid transform to other multiscale transforms for signals on graphs investigating which transforms work best on which types of graph signals in which signal processing tasks is beyond the scope of this work. However, in this section, we present a few illustrative numerical denoising and compression experiments with our proposed pyramid transform and other multiscale transforms for graph signals.

First, one qualitative difference between the proposed transform and many other multiscale transforms for graph signals is that our transform outputs both a multiresolution of graphs and a graph signal multiresolution residing on that graph multiresolution. While less important for numerical graph signal processing and machine learning tasks, this feature is especially beneficial for multiscale visualization of graph signals (e.g., the sequence of coarse approximations in the top row of Figure 8). A few other transforms such as the critically-sampled spline wavelet filter banks [9] share this feature, but many others such as the spectral graph wavelets [5] and the spatial graph wavelets [3] do not yield such a sequence of coarse approximations at different resolution levels.

Second, the dictionary atoms we construct tend to be jointly localized in the vertex domain and the graph spectral domain. The hope is that such atoms are able to sparsely represent various classes of graph signals, such as those that are piecewise smooth with discontinuities. Unfortunately, there is little theory to date relating mathematical classes of graph signals to the sparsity of particular transform coefficients. One initial exploration into this line of investigation for the case of spectral graph wavelets is presented in [55]. We conduct a few numerical experiments below to explore the decay of the transform coefficients for different types of signals.

1) *Denoising*: We start by repeating the denoising experiment of [56, Section VI.D] on the Minnesota road network of [52].¹⁰ We add white Gaussian noise with varying standard

¹⁰To replicate the experiment with our transform, we leveraged the MATLAB code of [56], which is publicly available at <http://tanaka.msp-lab.org/software>.

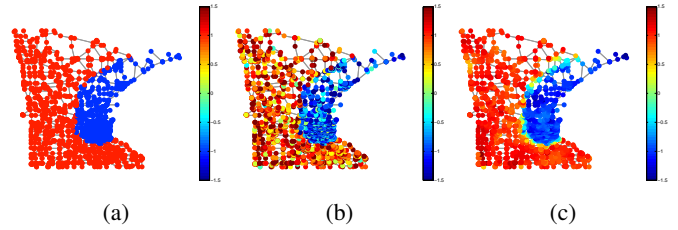


Fig. 10. Denoising example. (a) Piecewise constant signal on the Minnesota graph. (b) Noisy observation with $\sigma = \frac{1}{2}$. (c) Denoised signal reconstructed after hard thresholding the prediction errors of a two-level pyramid transform.

TABLE I
DENOISING RESULTS ON THE MINNESOTA TRAFFIC GRAPH: SNR (dB)

σ	noisy	1-level pyramid	2-level pyramid	reference range
1/32	30.10	34.81	34.30	31.44 - 35.08
1/16	24.08	28.60	27.35	25.61 - 29.34
1/8	18.06	22.84	20.34	19.97 - 23.17
1/4	12.04	17.00	15.85	14.19 - 17.63
1/2	6.02	12.77	13.51	8.50 - 12.31
1	0.00	7.96	10.46	2.63 - 8.82
Redundancy	-	1.50	1.76	1.00 - 4.00

deviations to the piecewise-constant reference signal from [19] shown in Figure 10(a). As in [56], we compute the pyramid transform coefficients, hard threshold the prediction errors at a threshold of 3σ , and reconstruct a denoised signal (shown in Figure 10(c) for the case of $\sigma = \frac{1}{2}$) from the coarse approximation coefficients and the thresholded prediction error coefficients. We use the same parameters as in Figure 7 to construct the graph multiresolution. We take the graph spectral filters to be $\hat{h}(\lambda_\ell) := \frac{1}{1+\lambda_\ell}$, and take $\epsilon = 0.005$ to form the regularized Laplacian for interpolation. In Table I, we compare the denoising results for a one-level and two-level transform to a reference range of results attained by different transforms in [56], including variants of the biorthogonal graph filter banks of [57], the oversampled graph filter banks of [56], and the spectral graph wavelet transform of [5]. Qualitatively comparing the denoised signal in Figure 10(c) to its counterparts in [56, Fig. 15], it appears that our denoised signal is smoother with respect to the graph structure, with more of the errors clustered around the discontinuity in the piecewise-constant signal.

2) *Compression*: Next, we repeat the compression experiment of Figure 9 on three different types of graph signals. First, we consider a piecewise-smooth signal with discontinuities. We compose the signal in Figure 11(a) on the 500 vertex random sensor network from Figure 4 by segmenting the graph into four strips and restricting two different polynomials of the coordinates to the different sections of the graph. The first polynomial $0.5 - 2\bar{x}$ is restricted to the first and third diagonal strips, counting from the upper right, and the second polynomial $0.5 + \bar{x}^2 + \bar{y}^2$ is restricted to the second and fourth strips (the latter of which is the lower left corner), where (\bar{x}, \bar{y}) are the physical coordinates of each vertex. We sum the two components to form the signal.

The second signal, shown in Figure 12(a), is from [58, Fig. 16(c)]. It is formed by generating four different random signals on the graph, filtering them with different bandpass filters, restricting each of the bandpass signals to different regions of the graph, and then summing them. The first three regions are

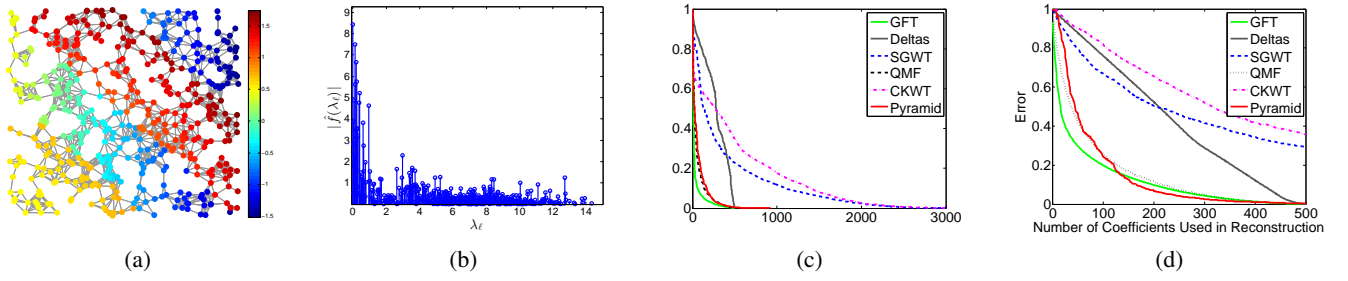


Fig. 11. Compression example I. (a) A piecewise-smooth graph signal on a random sensor network with 500 vertices. (b) The same signal in the graph spectral domain. (c) The normalized sorted magnitudes of the transform coefficients. There are 500 coefficients each for the deltas, graph Fourier transform, and QMF filter bank, 922 coefficients for the pyramid transform, and 3000 coefficients each for the spectral graph wavelets and spatial graph wavelets. (d) The reconstruction errors $\frac{\|f_{\text{reconstruction}} - f\|_2}{\|f\|_2}$ resulting from hard thresholding each transform's coefficients, as a function of the number of coefficients used in the reconstruction.

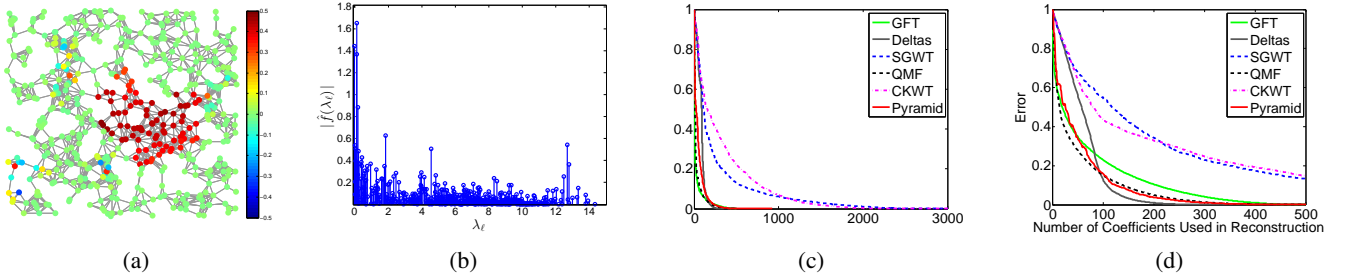


Fig. 12. Compression example II. (a) A signal comprised of four different bandpass signals restricted to four different parts of the graph. (b) The same signal in the graph spectral domain. (c) The normalized sorted magnitudes of the transform coefficients. (d) The reconstruction errors $\frac{\|f_{\text{reconstruction}} - f\|_2}{\|f\|_2}$.

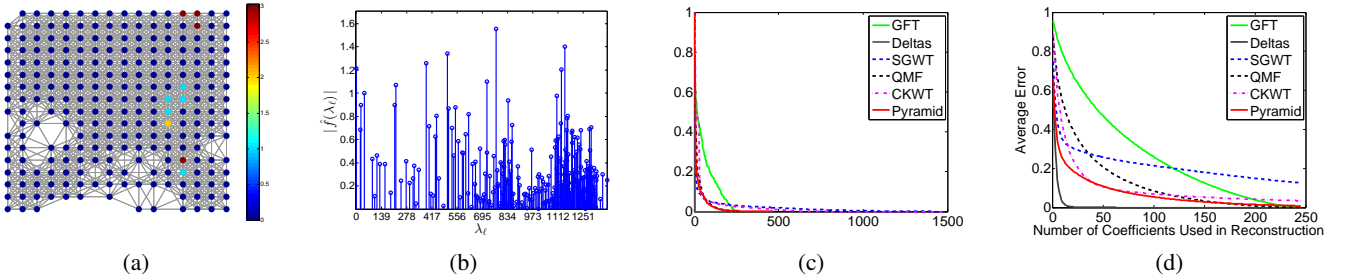


Fig. 13. Compression example III. (a) A signal from the Flickr data set of [59], [60]. (b) The same signal in the graph spectral domain. (c) The normalized sorted magnitudes of the transform coefficients of the same signal. (d) The reconstruction errors are averaged over 500 signals from the data set.

4-hop neighborhoods of different center vertices, and the last region contains all vertices in the graph that are not in one of the first three regions. We can see for example that the smooth component in the center of the graph was filtered by a lowpass kernel in the graph spectral domain.

The third class of signals we examine are from the Flickr data set, used previously in [59] and [60]. Each signal in the data set represents the number of distinct Flickr users that took photos in 10 x 10 meter regions around Trafalgar Square in London on a specific day. Like [60], we use a subgraph with 245 vertices, each one representing the centroid of a given 10 x 10 meter region. Note that while the signals from Figures 11(a) and 12(a) are sparser in the graph spectral domain than the vertex domain, this Flickr data set contains signals that are much sparser in the vertex domain than the graph spectral domain. Figures 13(a) and 13(b) represent one example signal from the database in both the vertex and graph spectral domains.

For all three compression experiments in this section, we

use a three-level pyramid transform, take $\epsilon = 0.005$ to form the regularized Laplacian for interpolation, and take the graph spectral filters to be lowpass Meyer kernels $\hat{h}^{(j)}(\lambda_\ell)$ whose values are equal to 1 on the interval $[0, \frac{\lambda_{\max}^{(j)}}{2}]$, smoothly decay to $\frac{\sqrt{2}}{2}$ at $\frac{3\lambda_{\max}^{(j)}}{4}$, and reach 0 by $\lambda_{\max}^{(j)}$, where the superscript j denotes the level of the pyramid. For comparison, we use (i) the graph Fourier transform (GFT); (ii) the basis of Kronecker deltas (i.e., compress the signal values directly); (iii) the spectral graph wavelet transform (SGWT) [5] with five wavelet scales plus the scaling functions for a total redundancy of 6; (iv) the spatial graph wavelets (CKWT) [3] with wavelet functions based on the renormalized one-sided Mexican hat wavelet, also with five wavelet scales and concatenated with the dictionary of Kronecker deltas (scale 0 dictionary atoms)¹¹ for a total redundancy of 6; and the graph-QMF filter bank

¹¹The spatial graph wavelets all have mean zero, so we add the dictionary of deltas to ensure that the dictionary analysis operator is full rank.

transform (QMF) [19].¹² Note that the graph-QMF transform only uses the adjacency matrix structure, and not the weights of the graphs.

Figures 11(c), 12(c), and 13(c) show the decay of the magnitudes of the transform coefficients for the signals shown in Figures 11(a), 12(a), and 13(a), respectively. In these plots, for each transform and signal pair, we normalize the sorted magnitudes by dividing every transform coefficient of that pair by the largest magnitude coefficient of that pair. For this compression experiment, we keep the M coefficients with the largest magnitudes, set the rest to 0, and reconstruct the signal from the thresholded transform coefficients. For a fair comparison, for all transforms, we perform least squares reconstruction by applying the pseudoinverse of the analysis operator to the thresholded coefficients. In Figures 11(d), 12(d), and 13(d), we show the reconstruction errors $\frac{\|\mathbf{f}_{\text{reconstruction}} - \mathbf{f}\|_2}{\|\mathbf{f}\|_2}$ for varying values of M , the number of coefficients used in the reconstruction. Note that the errors in 13(d) are averaged over 500 different signals from the data set. We see that our proposed pyramid transform works reasonably well for compression for all three signal models.

VII. IMPLEMENTATION APPROXIMATIONS AND COMPUTATIONAL COMPLEXITY

In their seminal paper on the Laplacian pyramid, Burt and Adelson [15] write, “The coding scheme outlined above will be practical only if the required filtering computations can be performed with an efficient algorithm.” In the same spirit, we mention here some approximation techniques to improve the computational efficiency of the proposed transform.

As described analytically, the transform relies on repeated eigendecompositions of the graph Laplacian for filtering, downsampling, and graph reduction. However, for large graphs (with tens of thousands or more vertices), computing full eigendecompositions is computationally expensive and may not be possible. This is because general-purpose routines for computing the full eigendecomposition have computational complexities of $O(N^3)$. Accordingly, two common themes running through this section are (i) to avoid computing full eigendecompositions, and (ii) to take advantage of (and preserve) the sparsity of the graph Laplacian whenever possible.

A. Power Method for Computing the Largest Eigenvector

To compute the largest eigenvector of an $N \times N$ graph Laplacian, we can use the *power method*, which generates a sequence of vectors $\{\mathbf{x}^{(k)}\}_{k=0,1,\dots}$ through the following recursion:

$$\mathbf{x}^{(k)} = \frac{\mathcal{L}\mathbf{x}^{(k-1)}}{\|\mathcal{L}\mathbf{x}^{(k-1)}\|_2}. \quad (21)$$

If $\lambda_{\max} > \lambda_{N-2}$ and $\langle \mathbf{x}^{(0)}, \mathbf{u}_{\max} \rangle \neq 0$, then the sequence of vectors $\{\mathbf{x}^{(k)}\}_{k=0,1,\dots}$ generated by the power method converges to \mathbf{u}_{\max} [61, Chapter 8.2.1]. Convergence is faster the smaller the ratio $\frac{\lambda_{N-2}}{\lambda_{\max}}$ is, and it may be accelerated by

subtracting a multiple of the identity matrix from \mathcal{L} to shift the spectrum [62, Chapter 4.1]. The computational cost of (21) is dominated by the $O(|\mathcal{E}|)$ matrix-vector multiplication by the matrix \mathcal{L} . In sparse graphs, $|\mathcal{E}|$ may grow linearly with N , making this an efficient method to compute the largest eigenvector. The condition $\langle \mathbf{x}^{(0)}, \mathbf{u}_{\max} \rangle \neq 0$ is satisfied by almost any $\mathbf{x}^{(0)}$ chosen at random [62, Chapter 5.1]. If the largest eigenvalue is not simple, the power method converges to a vector in the largest eigenspace, but this vector is not unique. For downsampling purposes, we only need to compute enough iterations to approximate the polarities of the largest eigenvector components, as opposed to their precise values.

B. Efficient Implementation of Kron Reduction and Storage of Reduced Graph Information

Full computation of the Kron-reduced Laplacian $\mathcal{K}(\mathcal{L}, \mathcal{V}_1)$ in (3) can be computationally expensive ($O(N^3)$ with naive methods that do not take advantage of the structure of $\mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1^c}$). One option is to not actually compute the entire reduced graph Laplacian, but rather to form an operator to do matrix-vector multiplication with the matrix being the Kron-reduced Laplacian (see, e.g., [63] or the function `MatCreateSchurComplement` in PETSc [64]). The drawback of doing so is that we cannot plot reduced graphs and coarser approximations as in Figure 7; however, such a method is well-suited for situations such as compression or regularization where we are only interested in computing transform coefficients.

In some cases, we may actually want to compute and store all of the graph Laplacians $\mathcal{L}^{(0)}, \mathcal{L}^{(1)}, \dots, \mathcal{L}^{(J-1)}$ as part of a multilevel transform. If we are repeatedly applying the transform to different signals on the same graph (e.g., a fixed transportation network), then the overhead of computing and storing this sequence of reduced graph Laplacians is shared across all of the applications of the transform. On the other hand, if the graph is completely different for every graph signal (e.g., some graph-based image processing methods), then efficient storage and reconstruction of the graph Laplacians is an important computational consideration. In either case, because the Schur complement is often dense even if \mathcal{L} is sparse, it can also be computationally expensive to store $\mathcal{L}^{\text{Kron-reduced}}$ for large graphs. The addition of a sparsification step after each Kron reduction reduces the storage requirements, as well as the complexity of subsequent computations. One interesting open question is whether we can avoid storing the entire sequence of Laplacians, but rather store the coarsest Laplacian and some additional side information, and still be able to recover the entire sequence in reverse order.

C. Approximate Graph Spectral Filtering

The Chebyshev polynomial approximation method, which is introduced in [5] and discussed further in [65], can be used to speed up graph spectral filtering operations. That method computes an approximately filtered signal $\tilde{\mathbf{H}}\mathbf{f}$ through repeated matrix-vector multiplication by the matrix \mathcal{L} at a computational cost of $O(K|\mathcal{E}|)$, where K is the order of the Chebyshev polynomial approximation. Therefore, for a sparse graph, it is far more efficient to compute $\tilde{\mathbf{H}}\mathbf{f}$ than $\mathbf{H}\mathbf{f}$.

¹²The MATLAB code for the graph-QMF filter bank is publicly available at http://biron.usc.edu/wiki/index.php/Graph_Filterbanks.

D. Computational Complexity

To summarize, we can break the computational complexity down into two parts: the first for computing the multiresolution of graphs and the second for analyzing signals on that multiresolution of graphs with the filtering and interpolation operations. As mentioned above, the overhead of computing and storing the multiresolution of graphs is often shared across applications of the transform to many different signals on the same graph.

The bottleneck for the signal analysis portion of the multiscale pyramid transform we proposed in Section VI is the interpolation step (16). Recent advances in solving such sparse, symmetric, diagonally dominant systems of equations provide methods to reduce the computational cost from $\mathcal{O}(|\mathcal{E}|N)$ down to $\mathcal{O}(|\mathcal{E}|\log^2 N)$, or even faster [14], [32], [66], [67]. These advances also enable the spectral sparsification step to be done in nearly linear time with respect to the number of edges in the original graph. To compute the full graph multiresolution for visualization, the bottleneck is the Kron reduction step (3). Here, we can again take advantage of methods for solving symmetric diagonally dominant systems of linear equations in nearly linear time in the number of edges. For example, we can use the LAMG solver of [32] to compute $\mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1^c}^{-1} \mathcal{L}_{\mathcal{V}_1^c, i}$, for each i in \mathcal{V}_1 . Both the time to setup the problem (create a sequence of coarse-grained Laplacians) and the time to solve the problem are empirically shown in [32] to be linear in the number of edges. Since we have to solve the problem for each column of $\mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1^c}$, the overall computational cost of the Kron reduction should be on the order $\mathcal{O}(|\mathcal{E}|N)$, but at least we only need to perform the setup portion once for $\mathcal{L}_{\mathcal{V}_1^c, \mathcal{V}_1^c}$.

VIII. CONCLUSION

Our main contribution in this work has been to present a new framework for a multiscale pyramid transform for graph signals that is centered around the four fundamental operations of graph downsampling, graph reduction, graph spectral filtering, and interpolation of graph signals. Our framework is modular, so we can easily substitute different choices for these fundamental operations (e.g., choose a different graph downsampling operator in the Laplacian pyramid of Figure 6). We hope this framework inspires new choices for and analyses of these fundamental operations. For example, since the presentation of an initial version of our work here, [68] has presented a new method for graph downsampling in the context of signal processing, and [69] has presented a new method for graph reduction that preserves the eigenvalues from one graph to the next in the graph multiresolution.

The interplay between the four operations is important and less well understood in the context of graph signal processing. Further empirical and theoretical investigations into how the joint choice of these operations affects properties such as the joint localization of the resulting dictionary atoms in the vertex-frequency space, as well as the sparsity of transform coefficients for different classes of graph signals is an important line of future work. Another natural offshoot of this work would be to combine these fundamental graph signal processing operations in different manners to generate other multiscale transforms such as critically sampled filter banks and lifting transforms.

REFERENCES

- [1] M. J. Faraji, "A Laplacian pyramid scheme in graph signal processing," Master's thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, Jun. 2011.
- [2] Shuman, D. I., S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, pp. 83–98, May 2013.
- [3] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Proc. IEEE INFOCOM*, vol. 3, Mar. 2003, pp. 1848–1857.
- [4] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 53–94, 2006.
- [5] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.
- [6] M. Jansen, G. P. Nason, and B. W. Silverman, "Multiscale methods for data on graphs and irregular multidimensional situations," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 71, no. 1, pp. 97–125, 2009.
- [7] S. K. Narang and A. Ortega, "Lifting based wavelet transforms on graphs," in *Proc. APSIPA ASC*, Sapporo, Japan, Oct. 2009, pp. 441–444.
- [8] —, "Local two-channel critically sampled filter-banks on graphs," in *Proc. Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 333–336.
- [9] V. N. Ekambaram, G. C. Fanti, B. Ayazifar, and K. Ramchandran, "Critically-sampled perfect-reconstruction spline-wavelet filter banks for graph signals," in *Proc. Glob. Conf. Signal Inf. Process.*, Austin, TX, Dec. 2013, pp. 475–478.
- [10] M. Gavish, B. Nadler, and R. R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 367–374.
- [11] S.-H. Teng, "Coarsening, sampling, and smoothing: Elements of the multilevel method," in *Parallel Processing, IMA Volumes in Mathematics and its Applications*, 105. Springer Verlag, 1999, pp. 247–276.
- [12] A. Brandt, "Multiscale scientific computation: Review 2001," in *Multiscale and Multiresolution Methods: Theory and Applications, Lecture Notes in Comput. Sci.* 20, T. J. Barth, T. Chan, and R. Haimes, Eds. Springer-Verlag, 2001, pp. 3–96.
- [13] D. Ron, I. Safro, and A. Brandt, "Relaxation-based coarsening and multiscale graph organization," *Multiscale Model. Simul.*, vol. 9, no. 1, pp. 407–423, Sep. 2011.
- [14] N. K. Vishnoi, *Lx = b*, ser. Foundations and Trends in Computer Science. Now Publishers, 2013.
- [15] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, pp. 532–540, Apr. 1983.
- [16] R. Roth, "On the eigenvectors belonging to the minimum eigenvalue of an essentially nonnegative symmetric matrix with bipartite graph," *Linear Algebra Appl.*, vol. 118, pp. 1–10, Jun. 1989.
- [17] T. Bıyıkoglu, J. Leydold, and P. F. Stadler, *Laplacian Eigenvectors of Graphs*. Lecture Notes in Mathematics, vol. 1915, Springer, 2007.
- [18] S. K. Narang and A. Ortega, "Downsampling graphs using spectral theory," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, Prague, Czech Republic, May 2011, pp. 4208–4211.
- [19] —, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.
- [20] G. Shen and A. Ortega, "Tree-based wavelets for image coding: Orthogonalization and tree selection," in *Proc. Picture Coding Symp.*, Chicago, IL, May 2009, pp. 1–4.
- [21] B. Aspöval and J. R. Gilbert, "Graph coloring using eigenvalue decomposition," *SIAM J. Alg. Disc. Meth.*, vol. 5, no. 4, pp. 526–538, 1984.
- [22] U. von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
- [23] E. B. Davies, G. M. L. Gladwell, J. Leydold, and P. F. Stadler, "Discrete nodal domain theorems," *Lin. Algebra Appl.*, vol. 336, no. 1-3, pp. 51–60, Oct. 2001.
- [24] G. Berkolaiko, "A lower bound for nodal count on discrete and metric graphs," *Commun. Math. Phys.*, vol. 278, no. 3, pp. 803–819, 2008.
- [25] I. Oren, "Nodal domain counts and the chromatic number of graphs," *J. Phys. A: Math. Theor.*, vol. 40, no. 32, pp. 9825–9832, 2007.
- [26] S. T. Barnard and H. D. Simon, "Fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems," *Concurrency: Practice and Experience*, vol. 6, no. 2, pp. 101–117, Apr. 1994.

- [27] F. Dörfler and F. Bullo, "Kron reduction of graphs with applications to electrical networks," *IEEE Trans. Circuits and Systems-I: Regular Papers*, vol. 60, no. 1, pp. 150–163, Jan. 2013.
- [28] F. Zhang, *The Schur Complement and its Applications*. Springer, 2005.
- [29] D. J. Klein and M. Randić, "Resistance distance," *J. Math. Chem.*, vol. 12, pp. 81–95, 1993.
- [30] D. A. Spielman and N. Srivastava, "Graph sparsification by effective resistances," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1913–1926, 2011.
- [31] S. Lafon and A. B. Lee, "Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1393–1403, Sep. 2006.
- [32] O. E. Livne and A. Brandt, "Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver," *SIAM J. Sci. Comput.*, vol. 34, no. 4, pp. B499–B522, 2012.
- [33] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 359–392, 1998.
- [34] I. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors a multilevel approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1944–1957, Nov. 2007.
- [35] C. Walshaw, "The graph partitioning archive," <http://staffweb.cms.gre.ac.uk/~wc06/partition/>.
- [36] F. K. Chung, *Spectral Graph Theory*. Vol. 92 of the CBMS Regional Conference Series in Mathematics, AMS Bokstore, 1997.
- [37] N. J. Higham, *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008.
- [38] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [39] —, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.
- [40] I. Pesenson, "Variational splines and Paley-Wiener spaces on combinatorial graphs," *Constr. Approx.*, vol. 29, no. 1, pp. 1–21, Feb. 2009.
- [41] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and semi-supervised learning on large graphs," in *Learn. Theory*, ser. Lect. Notes Comp. Sci. Springer-Verlag, 2004, pp. 624–638.
- [42] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Proc. Ann. Conf. Comp. Learn. Theory*, ser. Lect. Notes Comp. Sci., B. Schölkopf and M. Warmuth, Eds. Springer, 2003, pp. 144–158.
- [43] D. Zhou and B. Schölkopf, "Regularization on discrete spaces," in *Pattern Recogn.*, ser. Lect. Notes Comp. Sci., W. G. Kropatsch, R. Sablatnig, and A. Hanbury, Eds. Springer, 2005, vol. 3663, pp. 361–368.
- [44] X. Zhu and Z. Ghahramani, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Int. Conf. Mach. Learn.*, Washington, D.C., Aug. 2003, pp. 912–919.
- [45] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Adv. Neural Inf. Process. Syst.*, S. Thrun, L. Saul, and B. Schölkopf, Eds., vol. 16. MIT Press, 2004, pp. 321–328.
- [46] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning*. MIT Press, 2006.
- [47] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, Vancouver, BC, May 2013, pp. 5445–5449.
- [48] —, "Localized iterative methods for interpolation of graph structured data," in *Proc. IEEE Glob. Conf. Signal and Inform. Process.*, Austin, TX, Dec. 2013, pp. 491–494.
- [49] M. N. Do and M. Vetterli, "Frame reconstruction of the Laplacian pyramid," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, vol. 6, Salt Lake City, Utah, May 2001, pp. 3641–3644.
- [50] —, "Framing pyramids," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2329–2342, Sep. 2003.
- [51] O. Christensen, *Frames and Bases*. Birkhauser, 2008.
- [52] D. Gleich, "The MatlabBGL Matlab library," http://www.cs.purdue.edu/homes/dgleich/packages/matlab_bgl/index.html.
- [53] Stanford University Computer Graphics Laboratory, "The Stanford 3D Scanning Repository," <http://graphics.stanford.edu/data/3Dscanrep/>.
- [54] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.
- [55] B. Ricaud, D. I. Shuman, and P. Vandergheynst, "On the sparsity of wavelet coefficients for signals on graphs," in *SPIE Wavelets and Sparsity*, San Diego, California, Aug. 2013.
- [56] A. Sakiyama and Y. Tanaka, "Oversampled graph Laplacian matrix for graph filter banks," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6425–6437, Dec. 2014.
- [57] S. K. Narang and A. Ortega, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4673–4685, Oct. 2013.
- [58] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, Feb. 2015.
- [59] X. Dong, A. Ortega, P. Frossard, and P. Vandergheynst, "Inference of mobility patterns via spectral graph wavelets," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, Vancouver, Canada, May 2013, pp. 3118–3122.
- [60] D. Thanou, D. I. Shuman, and P. Frossard, "Learning parametric dictionaries for signals on graphs," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3849–3862, Aug. 2014.
- [61] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.
- [62] D. S. Watkins, *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods*. Society for Industrial and Applied Mathematics, 2007.
- [63] Y. Saad, *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.
- [64] S. Balay, J. Brown, K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang, "PETSc Web page," 2013, <http://www.mcs.anl.gov/petsc>.
- [65] D. I. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev polynomial approximation for distributed signal processing," in *Proc. Int. Conf. Distr. Comput. Sensor Sys.*, Barcelona, Spain, Jun. 2011.
- [66] I. Koutis, G. L. Miller, and R. Peng, "Approaching optimality for solving SDD linear systems," in *Proc. IEEE Symp. Found. Comp. Sci.*, Las Vegas, NV, Oct. 2010, pp. 235–244.
- [67] D. A. Spielman, "Laplacian linear equations, graph sparsification, local clustering, low-stretch trees, etc." <http://www.cs.yale.edu/homes/spielman/precon/precon.html>.
- [68] H. Q. Nguyen and M. N. Do, "Downsampling of signals on graphs via maximum spanning trees," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 182–191, Jan. 2015.
- [69] P. Liu, X. Wang, and Y. Gu, "Coarsening graph signal with spectral invariance," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, Florence, Italy, May 2014, pp. 1070–1074.